

UNIVERSITY OF ILLINOIS AT
CHICAGO
801 SO. MORGAN
CHICAGO, IL 60607



This book is the gift of

Dreyfus Fund

UNIVERSITY of ILLINOIS

Q

I

A773

V.5

1952

PER

AUSTRALIAN JOURNAL
OF
SCIENTIFIC RESEARCH

SERIES A
PHYSICAL SCIENCES

VOLUME 5

MELBOURNE

1952

AUSTRALIAN JOURNAL OF SCIENTIFIC RESEARCH

Published by the Council for Scientific and Industrial Research
Organization and the Australian National Research Council

Issued in two Series :

Series A – Physical Sciences

(Four issues each year appearing March, June, September, December)

Series B – Biological Sciences

(Four issues each year appearing February, May, August, November)

Price : Each series 30/- per annum, separate issues 7/6 each

EDITORIAL BOARD

Chairman and Editor : N. S. Noble

Members : Professor Sir Macfarlane Burnet

Professor E. J. Hartung

Professor L. H. Martin

Professor J. G. Wood

All enquiries and manuscripts should be forwarded to :

The Editor,

Australian Journal of Scientific Research,

Commonwealth Scientific and Industrial Research Organization,

314 Albert Street, East Melbourne, C.2, Victoria.

Printed by Australasian Medical Publishing Co. Ltd., Sydney.

CONTENTS

NUMBER 1, MARCH 1952

	PAGE
Approximations in Transient Surface Heating. By J. C. Jaeger	1
The Persistence of Meteor Trails. By L. G. H. Huxley	10
Radio-Frequency Radiation from the Constellation of Cygnus. By J. H. Piddington and H. C. Minnett	17
The Position and Movement on the Solar Disk of Sources of Radiation at a Frequency of 97 Mc/s. III. Outbursts. By Ruby Payne-Scott and A. G. Little	32
The High Latitude East-West Asymmetry of Cosmic Rays. By D. W. P. Burbury and K. B. Fenton	47
The Growth of Cloud Drops by Condensation. I. General Characteristics. By P. Squires	59
The Influence of Meteorological Conditions on Rainmaking in the Sydney Area. By E. J. Smith	87
The Stopping Power of Ilford C ₂ Emulsion. By A. J. Dyer	104
Influence of the Volume Fractions of the Phases on the Deformation of ($\alpha + \beta$) Brass. By L. M. Clarebrough and G. R. Perger	114
The Anisotropy of Electrical Resistivity of Cold-Drawn Wires of Some Cubic Metals and Alloys. By T. Broom and W. K. Clothier	119
On the Anisotropy of Electrical Resistivity of Deformed Cubic Metals and Alloys. By T. Broom	128
The Dielectric Properties of Systems Containing Straight Polar Chains. By R. A. Sack	135
The Viscosity and Rigidity of Gelatin in Concentrated Aqueous Systems. I. Viscosity. By C. W. N. Cumper and A. E. Alexander	146
The Viscosity and Rigidity of Gelatin in Concentrated Aqueous Systems. II. Rigidity. By C. W. N. Cumper and A. E. Alexander	153
Oleic Acid Monolayers on Concentrated Salt Solutions. By A. R. Gilby and the Late E. Heymann	160
The Magnetic Properties and Chemical Structures of Solid Haemins. By W. A. Rawlinson and P. B. Seutt	173
The Surface Chemistry of Proteins. III. The Rigidity of Adsorbed Films at an Air-Water Interface. By C. W. N. Cumper and A. E. Alexander	189
The Production of Intermediates in the Manufacture of Atebrin. By W. Davies, N. H. Olver, and B. W. Wilson	198
The Demethylation of Methoxyacridones. By G. K. Hughes, N. K. Matheson, A. T. Norman, and E. Ritchie	206

Short Communications

The Calculation of the Probable Error of Determinations of the Lunar Daily Harmonic Component Variations in Geophysical Data: A Correction. By Sydney Chapman	218
The Velocity of Sound in Gases at Low Pressures. By R. L. Abbey	223

NUMBER 2, JUNE 1952

	PAGE
The Bending of a Semi-infinite Strip. By R. C. T. Smith ..	227
Curve-Fitting by the Method of Grouping. By P. G. Guest ..	238
The Scattering of Fast Electrons and Positrons by Heavy Elements. By E. M. Gunnersen	258
The Distribution of the Discrete Sources of Cosmic Radio Radiation. By B. Y. Mills	266
On the Hysteresis of Adsorption on Solid Surfaces. By R. G. Wylie	288
A Study of the Annual Soil Temperature Wave. By E. S. West ..	303
The Rate of Natural Evaporation from Shallow Ponds. By J. Ferguson	315
The Spontaneous Emulsification of Mixtures of Oleic Acid and Paraffin Oil in Alkaline Solutions. By W. W. Mansfield	331
Conjugation Energy. By R. D. Brown	339
Prediction Curves for Counter-Current Separations: Application to Lignin Aldehydes. By D. E. Bland, W. E. Hillis, and E. J. Williams	346
Sulphonamides. III. Disulphonyl Derivatives of some <i>N</i> -Heterocyclic Amines. By S. J. Angyal, W. O. Morris, and W. K. Warburton	368
Sulphonamides. IV. The Reaction of <i>N</i> -Heterocyclic Amines with Sulphonyl Halides. By S. J. Angyal	374
The Chemistry of Eucalypt Kinos. II. Aromadendrin, Kaempferol, and Ellagic Acid. By W. E. Hillis	379
Alkaloids of the Australian Rutaceae: <i>Pentaceras australis</i> Hook. F. I. Isolation of the Alkaloids and Identification of Canthin-6-one. By H. F. Haynes, Eva R. Nelson, and J. R. Price	387
Alkaloids of the Australian Rutaceae: <i>Evodia xanthoxyloides</i> F. Muell. II. Isolation of the Alkaloids from the Leaves. By G. K. Hughes, K. G. Neill, and E. Ritchie	401
Alkaloids of the Australian Rutaceae: <i>Evodia xanthoxyloides</i> F. Muell. III. The Structures of the Coloured Alkaloids, Evoxanthidine, Xanthevodine, and Xanthoxoline. By J. R. Cannon, G. K. Hughes, K. G. Neill, and E. Ritchie	406
The Chemical Constituents of Australian <i>Flindersia</i> Species. III. The Alkaloids of <i>Flindersia collina</i> Bail. By F. A. L. Auet, P. T. Gilham, P. Gow, G. K. Hughes, and E. Ritchie	412

Short Communications

The Chemical Constituents of Australian <i>Flindersia</i> Species. IV. The Constituents of <i>Flindersia bourjotiana</i> F. Muell. By J. R. Cannon, G. K. Hughes, J. R. Price, and E. Ritchie	420
---	-----

NUMBER 3, SEPTEMBER 1952

	PAGE
A Simplified Form of the Relativistic Electromagnetic Equations. By N. W. Taylor	423
The Optimum Space-Charge-Controlled Focus of an Electron Beam. By D. L. Hollway	430
A Preliminary Survey of 1420 Mc/s. Line Emission from Galactic Hydrogen. By W. N. Christiansen and J. V. Hindman	437
The Positions of Six Discrete Sources of Cosmic Radio Radiation. By B. Y. Mills.. .. .	456
Ionospheric Measurements at Oblique Incidence over Eastern Australia. By M. Strohfeldt, R. W. E. McNicol, and G. de V. Gipps	464
The Growth of Cloud Drops by Condensation. II. The Formation of Large Cloud Drops. By P. Squires	473
Tensammetry: A Method of Investigating Surface Phenomena by A.C. Current Measurements. By B. Breyer and S. Hacobian	500
Isotope Effect in the Reaction of ^{14}C -Formaldehyde with Dimedone. By A. M. Downes	521
Liquid-Vapour Equilibria. III. The Systems Benzene- <i>n</i> -Heptane, <i>n</i> -Hexane-Chlorobenzene, and <i>cyclo</i> Hexane-Nitrobenzene. By I. Brown	530
An Investigation of Errors in the Amperometric and Starch Indicator Methods for the Titration of Millinormal Solutions of Iodine and Thiosulphate. By J. H. Bradbury and A. N. Hambly	541
The Viscosity of Urea Solutions of Keratin. By E. F. Woods	555
Alkaloids of the Australian Rutaceae: <i>Pentaceras australis</i> Hook. F. II. Identification of 5-Methoxycanthinone. By Eva R. Nelson and J. R. Price	563
<i>Short Communications</i>	
Angular Distribution of α -Particles from the $\text{Li}^7(p,\alpha)\text{He}^4$ Reaction at Low Proton Bombarding Energies. By F. Hirst	570
Kinetic Treatment Appropriate to Rapid Reactions involving Two Con- secutive Second Order Steps. By H. G. Higgins and E. J. Williams	572
The Crystal Structure of Potassium Hexachlororuthenate(IV). By C. S. Adams and D. P. Mellor	577
Alkaloids of the Australian Rutaceae: <i>Glycosmis pentaphylla</i> (Retz.) Correa. By A. W. McKenzie and J. R. Price	579

NUMBER 4, DECEMBER 1952

	PAGE
The Absolute Measurement of High-Energy Radiation. By B. M. Spicer	581
The Absorption and Reflection of Microwave Radiation by a Mercury-Vapour Discharge. By H. A. Prime	592
The Microwave Admittance of a Mercury-Vapour Discharge. By H. A. Prime	607
The Properties of Liquid Phase Embryos at the Intersections of Plane Solid Surfaces. By R. G. Wylie	618
The Condensation of a Vapour at a Crystalline Surface. By R. G. Wylie	628
The Frequency Dependence of the Dielectric Properties of Dipolar Substances. By J. J. O'Dwyer and R. A. Sack	647
Temperature Dependence of the Dielectric Properties of Long-Chain Aliphatic Alcohols in the Solid State. By J. S. Dryden	661
Dielectric Absorption and D.C. Conductivity in <i>n</i> -Primary Alcohols. By B. V. Hamon and R. J. Meakins	671
The Kinetics of Adsorption at Liquid Surfaces. By K. L. Sutherland ..	683
The Rôle of Surface Transport in the Stability and Breakdown of Foams. By W. E. Ewers and K. L. Sutherland	697
Thiohydantoins. I. The Preparation of Some 2-Thiohydantoins from Amino Acids and Acylamino Acids. By J. M. Swan	711
Thiohydantoins. II. Thiohydantoins Derived from Aspartic and Glutamic Acids. By J. M. Swan	721
Thiohydantoins. III. Anhydride Intermediates in the Formation of 1-Acyl-2-thiohydantoins from Acylamino Acids. By J. M. Swan..	728
The Reaction of Amino Acids and Proteins with Diazonium Compounds. I. A Spectrophotometric Study of Azo-Derivatives of Histidine and Tyrosine. By H. G. Higgins and Dallas Fraser	736
New Flavones from <i>Pongamia pinnata</i> (L.) Merr. By L. Ramachandra Row	754
Colouring Matters of Australian Plants. II. Naphthoquinones from <i>Diospyros hebecarpa</i> A. Cunn. By R. G. Cooke and H. Dowd ..	760
Alkaloids of the Australian Rutaceae: <i>Pentaceras australis</i> Hook. F. III. Identification of 4-Methylthiocanthin-6-one. By Eva R. Nelson and J. R. Price	768

Short Communications

Cosmic Ray Asymmetry Studies at Hobart, Tasmania. By D. W. P. Burbury and A. C. McLaren	782
Attempts to Effect Exchange and Displacement Reactions with Cobalt(II) Mesoporphyrin. By N. Ashelford and D. P. Mellor	784
On <i>N</i> -(2,4-Dinitrophenyl)phthalimide. By A. L. Beckwith and J. Miller	786
Corrigenda	789
Index to Volume 5	791

APPROXIMATIONS IN TRANSIENT SURFACE HEATING

By J. C. JAEGER*

[Manuscript received August 28, 1951]

Summary

A number of exact solutions for the surface temperature of a semi-infinite solid which is heated over a region of its surface is given. These have applications to edge effects in surface induction heating, to the heating of the anode of an X-ray tube, and to heating by friction. The regions discussed are a half-plane, a strip, a square, a rectangle, and a circle. Temperatures within the solid and the way in which they approximate to those of a continuous point source are discussed. Holm's approximation (heating of the region bounded internally by a sphere) and heat generation over a spherical region are also studied.

I. INTRODUCTION

It often happens that a solid is heated for a short time by constant flux of heat over portion of its surface. Frequently, under practical conditions, the curvature of the surface may be neglected so that the problem may be formulated as that of the semi-infinite solid $z > 0$, initially at zero temperature, which is heated for time $t > 0$ by constant flux over portion of the plane $z = 0$, there being no loss of heat from the remainder of this surface. In some cases the region to which heat is supplied approximates to a geometrically simple form such as a strip, rectangle, square, or circle: in other cases the shape of the region is not known, but it will appear that useful information can be inferred from the results for the simple shapes mentioned.

Suppose the solid $z > 0$ is of density ρ , specific heat c , thermal conductivity k , and diffusivity $\alpha = k/\rho c$. Then there are two fundamental results:

(i) If heat is supplied for $t > 0$ at the constant rate q per unit time per unit area over the whole of the plane $z = 0$, the temperature at depth z and time t is (Carslaw and Jaeger 1947, Section 25 (6))

$$\frac{2q(\alpha t)^{\frac{1}{2}}}{k} \operatorname{ierfc} \frac{z}{2(\alpha t)^{\frac{1}{2}}}, \quad \dots \dots \dots (1)$$

where

$$\operatorname{ierfc} x = \int_x^\infty \operatorname{erfc} \xi d\xi, \quad \operatorname{erfc} x = 1 - \operatorname{erf} x = \frac{2}{\pi^{\frac{1}{2}}} \int_x^\infty e^{-\xi^2} d\xi, \quad \dots \dots (2)$$

are tabulated functions. The surface temperature is

$$\frac{2q}{k} \left(\frac{\alpha t}{\pi} \right)^{\frac{1}{2}} \cdot \dots \dots \dots (3)$$

* University of Tasmania, Hobart.

(ii) If heat is supplied for $t > 0$ at the rate Q units per unit time at the origin, the temperature at time t at distance r from the origin is (Carslaw and Jaeger 1947, Section 104 (2))

$$\frac{Q}{2\pi kr} \operatorname{erfc} \frac{r}{2(\alpha t)^{1/2}} \dots\dots\dots (4)$$

These results, and all those in the sequel, may be obtained by integration from the solution for an instantaneous point source (Carslaw and Jaeger 1947, Section 102 (2)), viz. that if a quantity of heat Q is released at $t = 0$ at the origin, the temperature at distance r from it at time t is

$$\frac{Q}{4\pi r(\pi \alpha t)^{3/2}} e^{-r^2/4\alpha t} \dots\dots\dots (5)$$

The practical problems sort out immediately into two types: (i) those of heating over a large area for which (1) and (3) are the relevant approximations, and (ii) those of heating over a small area for which (4) is a useful approximation.

The first of these types arises in heating by a flat heating element and in high frequency induction heating. In these cases the surface temperature is usually calculated from (3) and the problem is that of determining the edge correction, that is, how much heat is lost from the region under the source by diffusion outwards. The variation of temperature over the surface for the cases in which heat is supplied over a half-plane, an infinite strip, and a square, respectively, is discussed from this point of view in Sections II and III. Heat supply over a circular area is studied in Section IV.

The second case referred to above, in which heat is supplied over a very small area, has been extensively studied in connexion with heat generation by friction and in the anode of an X-ray tube. In Section V the relation between the maximum and average surface temperatures for heated regions of various shapes is discussed and it is shown that, even for very elongated regions, the maximum temperature (which is usually relatively easy to calculate) provides a useful guide to the average temperature. Finally, in Section VI the temperature within a solid heated over a circular area is discussed, and it is found that at distances of the order of the diameter of the circle the effect is nearly the same as that given by (4) for a point source at the centre of the circle. Other exact solutions which may be used as approximations, namely, heat supply over a hemisphere and heat supply over the interior of a sphere, are discussed in Section VII.

II. HEAT SUPPLY OVER A HALF-PLANE

Suppose that heat is supplied at the constant rate q per unit time per unit area over the region $y = 0, -\infty < x < \infty$ of the surface $z = 0$. The temperature at time t at the point y of the surface, obtained by integrating the solution (Carslaw and Jaeger 1947, Section 103 (1)) for the instantaneous line source, is

$$\begin{aligned} \frac{q}{2\pi k} \int_0^t \frac{d\tau}{t-\tau} \int_{-\infty}^0 e^{-y^2/(4\alpha(t-\tau))} dy &= \frac{q\alpha^{1/2}}{2k\pi^{1/2}} \int_0^t \frac{d\tau}{(t-\tau)^{1/2}} \operatorname{erfc} \frac{y}{2\alpha^{1/2}(t-\tau)^{1/2}} \\ &= \frac{2q}{k} \left(\frac{\alpha t}{\pi} \right)^{1/2} f(y, t), \dots\dots\dots (6) \end{aligned}$$

$$\text{where } f(y, t) = \frac{1}{2} \operatorname{erfc} \frac{y}{2(\alpha t)^{\frac{1}{2}}} + \frac{y}{4(\pi \alpha t)^{\frac{1}{2}}} \operatorname{Ei} \left(-\frac{y^2}{4\alpha t} \right), \dots \dots \dots (7)$$

and $\operatorname{Ei}(-x)$ is the tabulated exponential integral defined by

$$\operatorname{Ei}(-x) = - \int_x^\infty \frac{e^{-u}}{u} du.$$

The surface temperature at any time is thus $f(y, t)$ times the value given by (1) for the case in which the solid is heated over the whole of its surface.

A very simple result can be obtained for the amount of heat which has crossed an infinite strip of unit width in the plane $y=0$ normal to the surface up to time t . This gives the heat which is effectively lost to the region under the heated surface. If v is the temperature at time t' at the point (x, y, z) , the required quantity is

$$\begin{aligned} & -k \int_0^t dt' \int_0^\infty \left[\frac{\partial v}{\partial y} \right]_{y=0} dz \\ &= -\frac{q}{2\pi} \int_0^t dt' \int_0^{t'} \frac{d\tau}{t' - \tau} \int_0^\infty dz \int_{-\infty}^0 \frac{y' dy'}{2\alpha(t' - \tau)} e^{-(z^2 + y'^2)/4\alpha(t' - \tau)} \\ &= \frac{2q\alpha^{\frac{1}{2}} t^{3/2}}{3\pi^{\frac{1}{2}}} \dots \dots \dots (8) \end{aligned}$$

III. HEAT SUPPLY OVER AN INFINITE STRIP AND OVER A SQUARE

Suppose that heat is supplied at the constant rate q per unit time per unit area over the infinite strip $-a < y < a$, $-\infty < x < \infty$ of the surface $z=0$. The surface temperature, which follows from (6), is

$$\frac{2q}{k} \left(\frac{\alpha t}{\pi} \right)^{\frac{1}{2}} F(\xi, T), \dots \dots \dots (9)$$

where

$$\begin{aligned} F(\xi, T) = & \frac{1}{2} \operatorname{erf} \frac{1+\xi}{2T^{\frac{1}{2}}} + \frac{1}{2} \operatorname{erf} \frac{1-\xi}{2T^{\frac{1}{2}}} \\ & - \frac{(1+\xi)}{4(\pi T)^{\frac{1}{2}}} \operatorname{Ei} \left\{ -\frac{(1+\xi)^2}{4T} \right\} - \frac{(1-\xi)}{4(\pi T)^{\frac{1}{2}}} \operatorname{Ei} \left\{ -\frac{(1-\xi)^2}{4T} \right\}, \dots (10) \end{aligned}$$

and

$$\xi = y/a, \quad T = \alpha t/a^2. \dots \dots \dots (11)$$

$F(\xi, T)$ then determines the temperature distribution across the surface of the solid. It is shown in Figure 1 for various values of T .

It is of some interest to calculate the corresponding result for heat supply over a square of side $2a$. In this case the distribution function corresponding to (10) for the direction through the centre parallel to a pair of sides is

$$\frac{1}{4T^{\frac{1}{2}}} \int_0^{2T^{\frac{1}{2}}} \operatorname{erf} \frac{1}{u} \left\{ \operatorname{erf} \frac{(1-\xi)}{u} + \operatorname{erf} \frac{(1+\xi)}{u} \right\} du. \dots \dots (12)$$

This function is shown by the dotted curves in Figure 1. It appears that the temperature distributions for the strip and square are indistinguishable up to about $T=0.1$, while for $T>1$ those for the square are substantially lower.

IV. HEAT SUPPLY OVER A CIRCLE

This case is of some theoretical interest and has been studied by many workers. Because there are often great uncertainties about the shape of a contact, the results for any not too irregular shape may in many cases be used equally well for comparison with experiment; further, it is found that the differences between the accurate values for shapes such as a circle of diameter $2a$ and a square of side $2a$ are small. Thus it seems reasonable to use the shape for which results can be obtained most simply. Usually the mathematics of the square or rectangle are less sophisticated than those of the circle, and for

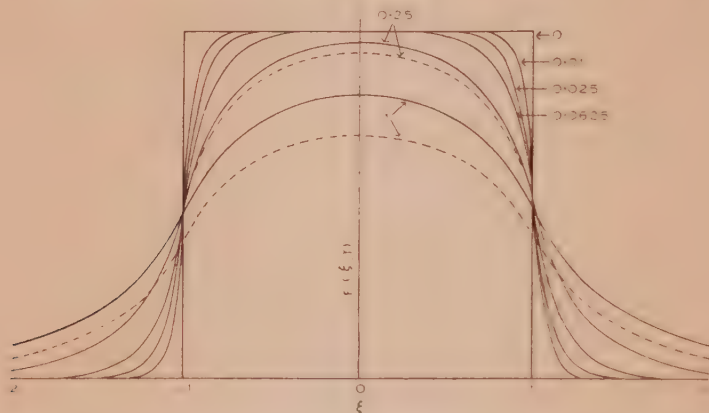


Fig. 1.—Temperature distribution across a strip (full lines) and a square (dotted lines) for various values of the time. The numbers on the curves are the values of $T = \alpha t/a^2$.

this reason these have been chosen by some authors, e.g. Blok (1937) and Jaeger (1943); but in some cases the circle leads to specially simple results, in particular at its centre, and for this reason has been used by Oosterkamp (1948*a*). Finally, because of the difficulties in the solution for a circle, Holm (1948) has introduced a physically simpler model; this, also, will be discussed below.

Here we give the complete solution for heat supply over a circle with formulae for important special cases. We suppose heat to be supplied for $t > 0$ over a circle of centre the origin and radius a in the surface of the solid $z > 0$. Then the temperature at the point $(0, 0, z)$ at time t is

$$\frac{2q(\alpha t)^{\frac{1}{2}}}{k} \left(\text{ierfc} \frac{z}{2(\alpha t)^{\frac{1}{2}}} - \text{ierfc} \frac{(z^2 + a^2)^{\frac{1}{2}}}{2(\alpha t)^{\frac{1}{2}}} \right) \quad (13)$$

This result, which follows by integration of (5), will be discussed later. Putting $z = 0$ gives the centre temperature

$$\frac{2aqT^{\frac{1}{2}}}{k\pi^{\frac{1}{2}}} \left(1 - \pi^{\frac{1}{2}} \text{ierfc} \frac{1}{2T^{\frac{1}{2}}} \right) \quad (14)$$

where

$$T = \alpha t/a^2. \quad (15)$$

The temperature at the point whose cylindrical coordinates are r, z may be put in the form*

$$\frac{aq}{2k} \int_0^\infty J_0(\lambda r) J_1(\lambda a) \left\{ e^{-\lambda z} \operatorname{erfc} \left(\frac{z}{2(\alpha t)^{\frac{1}{2}}} - \lambda(\alpha t)^{\frac{1}{2}} \right) - e^{\lambda z} \operatorname{erfc} \left(\frac{z}{2(\alpha t)^{\frac{1}{2}}} + \lambda(\alpha t)^{\frac{1}{2}} \right) \right\} \frac{d\lambda}{\lambda}, \quad \dots\dots\dots (16)$$

which generalizes the steady-state result (Carslaw and Jaeger 1947, Section 83 (7)) of Weber. The surface temperature, found by putting $z=0$ in (16), is

$$\frac{aq}{k} \int_0^\infty J_0(\lambda r) J_1(\lambda a) \operatorname{erf} \{ \lambda(\alpha t)^{\frac{1}{2}} \} \frac{d\lambda}{\lambda}. \quad \dots\dots\dots (17)$$

The centre temperature has already been found in an alternative form (14). The other interesting case is the temperature at the circumference $r=a$. In this case the result (17) may be transformed by using Watson (1944), Section 13.33 (1) and Section 5.4 (5) into

$$\frac{aq}{k} \left\{ \frac{2}{\pi} - \sum_{m=0}^\infty \frac{(-1)^m (2m)!}{\pi^{\frac{1}{2}} 2^{2m+1} m! [(m+1)!]^2 T^{m+\frac{1}{2}}} \right\}, \quad \dots\dots\dots (18)$$

and for small values of T has the value

$$\frac{aq}{k} \left\{ \frac{T^{\frac{1}{2}}}{\pi^{\frac{1}{2}}} - \frac{T}{2\pi} - \frac{T^2}{16\pi} - \dots \right\}. \quad \dots\dots\dots (19)$$

Another interesting quantity is the average temperature over the circle. This is

$$\frac{2q}{k} \int_0^\infty J_1^2(\lambda a) \operatorname{erf} [\lambda(\alpha t)^{\frac{1}{2}}] \frac{d\lambda}{\lambda^2}, \quad \dots\dots\dots (20)$$

which may be transformed into the series

$$\frac{2qa}{k} \left\{ \frac{4}{3\pi} - \sum_{m=0}^\infty \frac{(-1)^m (2m+2)!}{2^{2m+2} (m+2)! [(m+1)!]^2 (2m+1) \pi^{\frac{1}{2}} T^{m+\frac{1}{2}}} \right\}, \quad \dots\dots (21)$$

and for small values of T has the value

$$\frac{2qa}{k} \left\{ \frac{T^{\frac{1}{2}}}{\pi^{\frac{1}{2}}} - \frac{T}{\pi} + \frac{T^2}{8\pi} + \dots \right\}. \quad \dots\dots\dots (22)$$

The graphs of the centre, average,† and circumferential temperatures as functions of T are shown in curves I, II, and III of Figure 2. For large values of T they tend to aq/k , $8aq/3k\pi$, and $2aq/k\pi$, respectively.

V. AVERAGE TEMPERATURE

For the simple shapes discussed above it is usually relatively easy to calculate the centre temperature, which, of course, is also the maximum temperature. The other quantity of some interest is the average temperature; this not only

* The result follows after some reduction from Carslaw and Jaeger (1947), Section 103 (7), and Weber's integral (Watson (1944), Section 13.31).

† Oosterkamp (1948*b*) gives a curve for the average temperature which lies below that of Figure 2.

gives a measure of the temperature over the whole surface but, in the important case in which temperature is measured by using metals in contact as a thermocouple, it represents (ideally) the quantity actually observed.

In all cases the ratio of maximum to average temperature tends to unity as $t \rightarrow 0$, and it increases monotonically to its steady-state value in a manner shown for a circle in Figure 2.

If v_m and v_{av} are the maximum and average temperatures in the steady state, the ratio v_m/v_{av} has the value 1.186 for a square (Jaeger 1943) and the

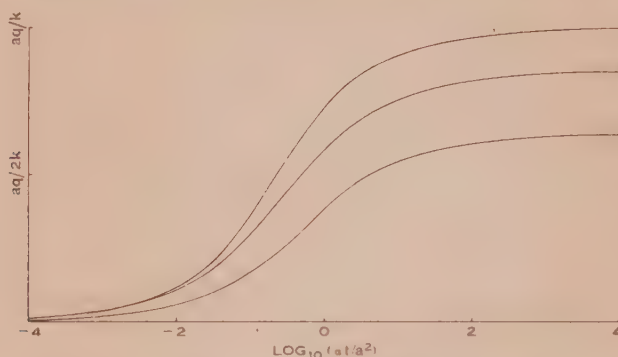


Fig. 2.—Centre, average, and circumferential temperature for heating over a circle of radius a : curves I, II, and III.

value $3\pi/8 = 1.178$ for a circle. For a rectangle of length $2l$ and width $2a$ the results are

$$v_m = \frac{2q}{k\pi} \left\{ a \sinh^{-1} \frac{l}{a} + l \sinh^{-1} \frac{a}{l} \right\}, \quad \dots \dots \dots (23)$$

$$v_{av} = \frac{2q}{k\pi a l} \left\{ l a^2 \sinh^{-1} \frac{l}{a} + a l^2 \sinh^{-1} \frac{a}{l} + \frac{1}{3} [a^3 + l^3 - (l^2 + a^2)^{3/2}] \right\}, \quad \dots (24)$$

so that for $l/a = 1, 2, 4, 10$, v_m/v_{av} has the values 1.186, 1.178, 1.159, and 1.133, respectively.

It appears that the ratio v_m/v_{av} varies very little with shape, even for quite elongated regions, and is never likely to be greater than 1.2. In view of the uncertainties in the physical assumptions it seems that a calculation of the maximum temperature will provide an adequate indication of surface temperature.

VI. TEMPERATURES AWAY FROM THE SOURCE OF HEAT

The way in which the temperature due to a circular disk source approximates to that due to a point source at its centre which emits heat at the same rate will now be discussed numerically. To do this we write the result (13) for the temperature at the point distant r from the origin along a perpendicular to the surface through the centre of the circle in the form

$$\frac{Q}{2\pi k r l} \left\{ 4\pi^2 T_1^4 \left(\operatorname{ierfc} \frac{1}{2T_1^{\frac{1}{2}}} - \operatorname{ierfc} \frac{(1 + \frac{r^2}{4l^2})^{\frac{1}{2}}}{2T_1^{\frac{1}{2}}} \right) \right\}, \quad \dots \dots \dots (25)$$

where

$$Q = \pi a^2 q, \quad T_1 = \alpha t / r^2, \quad \rho = r/a, \quad \dots \dots \dots (26)$$

and compare it with the corresponding result (4), viz.

$$\frac{Q}{2\pi k r} \operatorname{erfc} \frac{1}{2T_1^{1/2}}, \quad \dots \dots \dots (27)$$

The result (27) for the point source is shown in curve I of Figure 3 and the values of (25) for $\rho=2$ and $\rho=1$ in curves II and III, respectively. It appears that in this particular direction and at distances greater than a diameter the circular source behaves effectively as a point source. A similar result may be anticipated in other directions in which the exact calculation for the circle is more difficult.

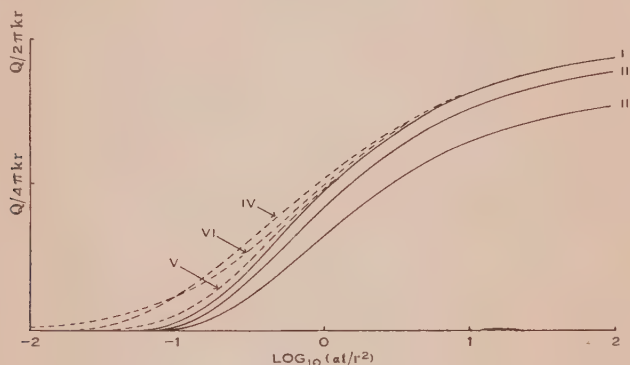


Fig. 3.—Temperature due to a point source, curve I; a disk source, curves II and III; Holm's approximation, curves IV and V; and a spherical source, curve VI.

VII. OTHER APPROXIMATIONS

There are two relatively simple problems on symmetrical radial flow in spherical coordinates which are of interest in the present context. The first of these considers the region bounded internally by the sphere $r=b$ with zero initial temperature and with heat supplied at the constant rate q' per unit time per unit area of the sphere. In this case the temperature at radius r and time t is

$$\frac{b^2 q'}{kr} \left\{ \operatorname{erfc} \frac{r-b}{2(\alpha t)^{1/2}} - \exp \left(\frac{r-b}{b} + \frac{\alpha t}{b^2} \right) \operatorname{erfc} \left(\frac{r-b}{2(\alpha t)^{1/2}} + \frac{(\alpha t)^{1/2}}{b} \right) \right\}. \quad \dots (28)$$

This has been used by Holm (1948) as an approximation to the solution of the problem of Section IV in the following way: he takes the total heat supply over a hemisphere to be equal to that over the circle of Section IV, so that $qa^2 = 2q'b^2$, also he chooses b so that the final temperature of the sphere is equal to the final average temperature over the circle which he takes to be $\pi a q / 4k$, leading to

$$b = 2a/\pi. \quad \dots \dots \dots (29)$$

The values of the temperatures for $r=a$ and $r=2a$ with these values of q' and b , and $Q = \pi q a^2$, are shown in curves IV and V of Figure 3. It appears that at distances greater than $2a$ Holm's result approximates very nearly to

that for the point source. It should be remarked that Holm deduces the final average temperature over the circle from physical considerations and his value is different from the arithmetic mean value of $8aq/3\pi k$ found in Section IV. If this value were used, it would lead to $b=3\pi a/16$ in place of (29).

The second simple problem in radial flow referred to above is that in which heat is supplied at a uniform rate of A_0 per unit time per unit volume over the spherical region $r < a$, the solid being initially at zero temperature. The temperature for $t > 0$ is

$$\frac{\alpha A_0 t}{k} \left\{ \frac{2a}{r} i^2 \operatorname{erfc} \frac{r-a}{2(\alpha t)^{\frac{1}{2}}} + \frac{2a}{r} i^2 \operatorname{erfc} \frac{r+a}{2(\alpha t)^{\frac{1}{2}}} - \frac{4(\alpha t)^{\frac{1}{2}}}{r} i^3 \operatorname{erfc} \frac{r-a}{2(\alpha t)^{\frac{1}{2}}} + \frac{4(\alpha t)^{\frac{1}{2}}}{r} i^3 \operatorname{erfc} \frac{r+a}{2(\alpha t)^{\frac{1}{2}}} \right\}, \quad \dots \dots \dots (30)$$

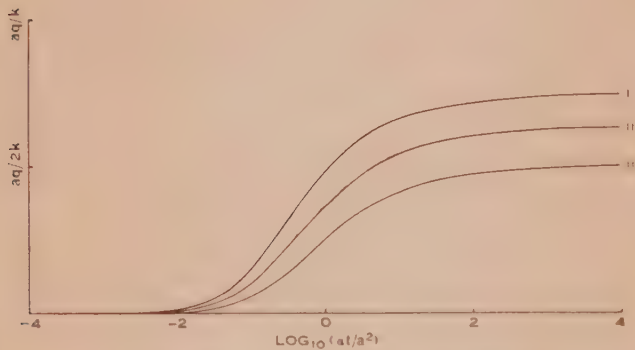


Fig. 4.—Centre, average, and circumferential temperatures for heating over a sphere of radius a : curves I, II, and III.

for $r > a$, and

$$\frac{\alpha A_0 t}{k} \left\{ 1 - \frac{2a}{r} i^2 \operatorname{erfc} \frac{a-r}{2(\alpha t)^{\frac{1}{2}}} + \frac{2a}{r} i^2 \operatorname{erfc} \frac{a+r}{2(\alpha t)^{\frac{1}{2}}} - \frac{4(\alpha t)^{\frac{1}{2}}}{r} i^3 \operatorname{erfc} \frac{a-r}{2(\alpha t)^{\frac{1}{2}}} + \frac{4(\alpha t)^{\frac{1}{2}}}{r} i^3 \operatorname{erfc} \frac{a+r}{2(\alpha t)^{\frac{1}{2}}} \right\}, \quad \dots (31)$$

for $r < a$, while for $r=0$ the result is

$$\frac{\alpha A_0 t}{k} - \frac{2A_0(\alpha t)^{\frac{1}{2}}a}{k} i \operatorname{erfc} \frac{a}{2(\alpha t)^{\frac{1}{2}}} - \frac{4\alpha t A_0}{k} i^2 \operatorname{erfc} \frac{a}{2(\alpha t)^{\frac{1}{2}}}, \quad \dots (32)$$

Equation (32) replaces Carslaw and Jaeger (1947), Section 129 (12), which is incorrect. The quantities $i^n \operatorname{erfc} x$ appearing in (30) to (32) are defined by (2) and

$$i^n \operatorname{erf} x = \int_x^\infty i^{n-1} \operatorname{erfc} \xi d\xi, \quad n=1, 2, \dots,$$

and are tabulated in Carslaw and Jaeger (1947). For comparison with the problem of Section IV we must choose the total rate of heat supply in the two cases to be the same, that is

$$4\pi a^3 A_0/3 = 2\pi a^2 q, \quad \dots \dots \dots (33)$$

Curves I and III of Figure 4 show the temperatures at $r=0$ and $r=a$, while curve II gives the average temperature over a circle of radius a through the origin. In these A_0 is given the values (33) so that they may be compared with the curves of Figure 2. The temperature at $r=a$ is also shown in curve VI of Figure 3 which indicates that, except for small values of the time, the temperature at $r=a$ is not very different from that at the same distance from a point source which emits heat at the same total rate.

It is of some interest to compare the results of Figures 2 and 4 which correspond to liberating heat at the same total rate over a circle and over a sphere of the same diameter. In problems on frictional heating it is frequently tacitly assumed that the heat is liberated in a surface, while in fact the effect takes place over some (unknown) volume surrounding this surface: thus the results of Figure 2 may be regarded as an upper limit while those of Figure 4 illustrate the effect of distributing the heat over a relatively large volume.

VIII. ACKNOWLEDGMENT

I am indebted to Mr. C. H. Johnson for assistance with the numerical work.

IX. REFERENCES

- BLOK, H. (1937).—"General Discussion on Lubrication." Vol. 2. p. 222. (Instn. Mech. Engrs.: London.)
CARSLAW, H. S., and JAEGER, J. C. (1947).—"Conduction of Heat in Solids." (Oxford Univ. Press.)
HOLM, R. (1948).—*J. Appl. Phys.* **19**: 361.
JAEGER, J. C. (1943).—*J. Roy. Soc. N.S.W.* **76**: 203.
OOSTERKAMP, W. J. (1948a).—*Philips Res. Rep.* **3**: 49.
OOSTERKAMP, W. J. (1948b).—*J. Appl. Phys.* **19**: 1180.
WATSON, G. N. (1944).—"Theory of Bessel Functions." 2nd Ed. (Cambridge Univ. Press.)

THE PERSISTENCE OF METEOR TRAILS

By L. G. H. HUXLEY*

[*Manuscript received September 4, 1951*]

Summary

The theory of ambipolar diffusion is applied to the spreading of meteor trails.

I. INTRODUCTION

A theoretical investigation of the expansion by diffusion in an unlimited atmosphere of a group of equal numbers of positive ions and electrons is of interest for the study both of the persistence of meteor trails and of irregularities in the ionosphere. In what follows it is shown how such an investigation may be carried out by application of the conventional theory of concentration cells (Abraham 1932, p. 118) and of the uniform positive column (Townsend 1928; Cobine 1941).

II. SYMBOLS; RELEVANT DATA CONCERNING THE MOTIONS OF ELECTRONS AND IONS IN GASES

The subscript 1 refers to positive ions and 2 to electrons.

Let

n_1, n_2 be the number of particles per unit volume,

K_1, K_2 diffusion coefficients,

W_1, W_2 drift velocities under electric field E ,

a_1, a_2 mobilities, i.e. $W_1 = a_1 E, W_2 = -a_2 E$,

k , Townsend's energy factor. k is a function of E/p .

$b = a_1/K_1$.

$N_1 = N_2 = N$ the number both of positive ions and of electrons in the whole group.

p is the pressure of the gas measured in mm. of mercury.

The following facts are known from studies of the motion of ions and electrons in gases (Cobine 1941; Townsend 1948; Huxley and Zaazou 1949),

$$b = a_1/K_1 = n_e e/P; \quad a_2/K_2 = b/k; \quad W_1 K_2 = k W_2 K_1; \quad a_1 K_2 = k a_2 K_1,$$

where $n_e e$ is the product of Avogadro's number (per unit volume) and the atomic charge and P is atmospheric pressure corresponding to n_e . Thus in electrostatic and c.g.s. units

$$n_e e = 1.22 \times 10^{10}; \quad P = 1.013 \times 10^6 \text{ dynes/cm.}^2$$

Thus

$$b = a_1/K_1 = k a_2/K_2 = 1.2 \times 10^4; \quad \text{also } a_1 = h/p,$$

* Department of Physics, University of Adelaide.

where h is a constant. The mobility of positive ions in air is approximately $a_1 = 3 \times 10^5/p = h/p$ in electrostatic and c.g.s. units, with p expressed in mm. of mercury. Consequently,

$$K_1 = a_1/b = 25/p.$$

The drift velocities W_2 of electrons are not, in general, simple functions of E/p , consequently in the relationship $W_2 = -a_2 E$, a_2 is not a constant independent of E . In fact, when $E/p < 1$ (E in V./cm.), $W_2 \propto (E/p)^{1/2}$ in air. It would be impracticable to use the rigorous dependence of W_2 upon E/p in what follows, and in practice, what is required is the order of magnitude of W_2 in order to assess which terms may be legitimately discarded as small and which must be retained. In air, when $E/p = 0.5$ (E in V./cm.), $W_2 = 9 \times 10^5$ cm./sec. (Nielson and Bradbury 1937). Thus, a crude value of a_2 (the electron mobility) within the range $0 < E/p < 1$ is $a_2 = 1.8 \times 10^6/p$ in these units. If E is expressed in electrostatic units, then $a_2 = 5 \times 10^8/p$. The ratio a_1/a_2 is therefore a small quantity of the order of magnitude 10^{-3} . The ratio of the diffusion coefficients $K_1/K_2 = a_1/ka_2$ is even smaller.

III. MATHEMATICAL TREATMENT

Consider a diffusing cloud of intermingled positive ions and electrons. Immediately after its formation, since the electrons in the group diffuse more rapidly from it than the positive ions, a positive space charge is established which gives rise to a field E which inhibits the further separation of positive ions and electrons by differential diffusion.

Provided the initial total number of ions and electrons N in the whole group is sufficiently large, the space charge $(n_1 - n_2)e$ can automatically adjust itself to control the mutual coherence of the two groups, electrons and positive ions.

Given the initial state, we require therefore the distributions in space of n_1 and n_2 at any subsequent time, as well as those of $(n_1 - n_2)$ and E . It is therefore necessary to solve the following equations:

$$\partial n_1 / \partial t = K_1 \nabla^2 n_1 - \text{div}(n_1 W_1) = K_1 \nabla^2 n_1 - a_1 \text{div}(n_1 E), \quad \dots \quad (1)$$

$$\partial n_2 / \partial t = K_2 \nabla^2 n_2 - \text{div}(n_2 W_2) = K_2 \nabla^2 n_2 + a_2 \text{div}(n_2 E), \quad \dots \quad (2)$$

$$\text{div } E = (n_1 - n_2)e/\epsilon_0. \quad \dots \quad (3)$$

Equations (1) and (2) are the equations of continuity and (3) Poisson's equation, ϵ_0 being the permittivity of free space in a rationalized system of electrical units. Multiply equation (1) by a_2 and equation (2) by a_1 and add the equations to obtain

$$\partial / \partial t (a_2 n_1 + a_1 n_2) = \nabla^2 (a_2 K_1 n_1 + a_1 K_2 n_2) - a_1 a_2 \text{div}(n_1 - n_2) E = 0. \quad \dots \quad (4)$$

In order to simplify this equation we note that in practice the difference $(n_1 - n_2)$ which provides the controlling space charge is extremely small compared with n_1 or n_2 . It is legitimate, therefore, to discard the term $a_1 a_2 \text{div}(n_1 - n_2) E$ in (4) and to write $n_1 = n_2 = n$ in the remaining terms. Thus from (4) we obtain

$$\partial n / \partial t = K \nabla^2 n, \quad \dots \quad (5)$$

where

$$K = (a_2 K_1 + a_1 K_2) / (a_1 + a_2). \quad \dots\dots\dots (6)$$

But, as explained in Section II, $a_1 \ll a_2$ and $a_1 K_2 / a_2 = k K_1$; the expression for K therefore is equivalent to

$$K = K_1(1+k) = K_2(1+1/k)a_1/a_2 \doteq 10^{-3} K_2. \quad \dots\dots\dots (7)$$

Equation (5) describes the coherent expansion of the group of intermingled positive ions and electrons. The apparent diffusion coefficient of the positive ions is increased by a relatively small factor $(1+k)$ and that of the electrons reduced by the factor $(1+1/k)a_1/a_2$. (The Earth's magnetic field may be ignored because $K_2 \gg K_1$ and $W_2 \gg W_1$ at all heights below 90 km.)

Derivation of Expression for E

Since $n_1 \doteq n_2 \doteq n$ equations (2) and (5) are equivalent; consequently,

$$\partial n / \partial t = \text{div} (K \text{ grad } n) = \text{div} (K_2 \text{ grad } n - n \mathbf{W}_2),$$

that is to say

$$n \mathbf{W}_2 = (K_2 - K) \text{ grad } n \doteq K_2 \text{ grad } n$$

since $K \doteq 10^{-3} K_2$.

Also

$$\mathbf{W}_2 = -a_2 \mathbf{E},$$

whence

$$\mathbf{E} = -(K_2/a_2) \text{ grad} (\log n) = -(k/b) \text{ grad} \log n. \quad \dots\dots\dots (8)$$

\mathbf{E} is therefore derivable from a potential $V = (k/b) \log n$ and may be found from the appropriate solution of equation (5).

Derivation of an Expression for $(n_1 - n_2)$

From equations (3) and (8) it follows that

$$(n_1 - n_2) = (\epsilon_0/e) \text{ div } \mathbf{E} = -(k\epsilon_0/be) \text{ div} (\text{grad} \log n). \quad \dots\dots\dots (9)$$

The three essential formulae are therefore

$$\partial n / \partial t = K \nabla^2 n, \quad \dots\dots\dots (10a)$$

$$\mathbf{E} = -(k/b) \text{ grad} (\log n), \quad \dots\dots\dots (10b)$$

$$(n_1 - n_2) = (\epsilon_0/e) \text{ div } \mathbf{E}, \quad \dots\dots\dots (10c)$$

with

$$K = K_1(1+k); \quad b = a_1/K_1 (= 1.2 \times 10^4 \text{ in e.s.u.}).$$

IV. APPLICATIONS

(a) Diffusion from a Group Strongly Localized at $t=0$

Let a group of N positive ions and an equal number of electrons be formed within a small volume at $t=0$ at the origin of coordinates. The solution of (10a) which is adapted to these initial conditions is

$$n = [N/(4\pi Kt)^{3/2}] \exp(-r^2/4Kt), \quad \dots\dots\dots (11)$$

which makes n zero everywhere except at $r=0$ when $t=0$, but at all times $t>0$,

$$4\pi \int_0^\infty nr^2 dr = N$$

(Carslaw and Jaeger 1947). According to (10b) and (11) the electric field in this expanding spherical cloud is

$$\mathbf{E} = \mathbf{r}/2Kt = [k/(k+1)](\mathbf{r}/2a_1t) = [k/(k+1)](\mathbf{r}p/2ht),$$

that is

$$(E/p) = [k/(k+1)](r/2ht), \dots\dots\dots (12)$$

where \mathbf{r} is a radial vector with $|\mathbf{r}| = r$, and $a_1 = h/p$. From (10c) and (12) it follows that

$$(n_1 - n_2) = 3k\varepsilon_0 p/[2(k+1)ht] = 3\varepsilon_0 E/er. \dots\dots\dots (13)$$

The following conclusions may be drawn from equations (11), (12), and (13) concerning the diffusing group:

- (i) The concentration difference $(n_1 - n_2)$ and the space charge $(n_1 - n_2)e$ are independent of r but are inversely proportional to t . This result is not true for distances r and times t which make n_1 as given by (11), less than the right-hand side of (13).
- Beyond this distance the space charge diminishes and \mathbf{E} does not suffice to retain electrons which diffuse beyond this point. The value of $r = r_c$ which makes n given by (11) equal to the right-hand side of (13) may be conveniently taken as representing the radius of the group at time t .
- (ii) The field \mathbf{E} and the space charge $(n_1 - n_2)e$ for $r < r_c$ are independent of the initial population N of the group.
- (iii) E/p is independent of the pressure of the gas.
- (iv) At a fixed distance r the concentration n attains its maximum value at time t given by

$$r^2 = 6Kt, \dots\dots\dots (14)$$

the value of $n_{max}(r)$ at this time t is $N/\left[\left(\frac{2}{3}\pi\bar{e}\right)^{3/2}r^3\right]$, where \bar{e} is the base of

natural logarithms. It is of interest to note that the mean square radius of the group is also given by (14); that is

$$\bar{r}^2 = 4\pi \int_0^\infty nr^4 dr = 6Kt \text{ (Townsend 1915, p. 85).}$$

The local maximum concentration $n_{max}(r)$ is therefore to be found at the distance $(\bar{r}^2)^{1/2}$ from the centre of diffusion, at any time t . Since $n_{max}(r)$ is inversely proportional to r^3 , it follows that the concentration n at distances greater than a specified distance $> r_{max}$ are always less than $n_{max}(r)$. When the constants in the expression for $n_{max}(r)$ are given their numerical values, we find

$$n_{max}(r)r^3 = N/13 \cdot 5 = 0 \cdot 074N. \dots\dots\dots (15)$$

(b) Diffusion from a Linear Concentration

We consider an initial linear concentration, such as a meteor trail, with N positive ions and the same number of electrons per unit length.

Equation (5) is now to be expressed in terms of the radial distance r from the axial concentration and the appropriate solution is (Carslaw and Jaeger 1947)

$$n = [N/(4\pi Kt)] \exp(-r^2/4Kt). \quad (16)$$

It follows from (10b) and (10c) that

$$E/p = [k/(k+1)](r/2ht), \quad (17)$$

as for the sphere (equation (12)); and

$$(n_1 - n_2) = p k \varepsilon_0 / [(k+1) h e t] = 2 \varepsilon_0 E / r e. \quad (18)$$

The conclusions (i), (ii), and (iii) of Section IV (a) are also valid for cylindrical diffusion.

The local maximum of concentration $n_{max}(r)$ at a specified radial distance r occurs at time t given by

$$r^2 = 4Kt - \bar{r}^2. \quad (19)$$

Whence, it follows from equation (16) that

$$\left. \begin{aligned} n_{max}(r)r^2 &= N/(\pi \bar{r}) = 0.117 N, \\ t &= r^2/4K = 0.029 [N/K n_{max}(r)]. \end{aligned} \right\} \quad (20)$$

V. THE ELECTRIC FIELD WITHIN A DIFFUSING COLUMN

Both for a cylinder and a sphere the electric field is to be obtained from the formula (12) which is the same as (17). Put $h = 3 \cdot 10^5$ (Section II); if k is large, $k/(k+1) \approx 1$. If $k = 1$ (its smallest possible value) $k/(k+1) = 0.5$. Thus

$$\frac{10^{-5}}{12} \frac{r}{t} \leq E/p \leq \frac{10^{-5}}{6} \frac{r}{t}; \quad E \text{ in e.s.u.}$$

Put $t = 1$ sec.; $r = 10$ cm., then

$$2.5 \times 10^{-5} \leq E/p \leq 5 \times 10^{-5}.$$

This is a small value of E/p and in it $k \approx 1$. Thus

$$E/p = 2.5 \times 10^{-5} \text{ (e.s.u.)} = 2.5 \times 10^{-3} \text{ (V./cm.)}.$$

If $r = 1$ cm., $t = 10^{-2}$ sec., $E/p = 2.5 \times 10^{-2}$ (V./cm.). The difference in concentrations is (formulae (13) and (18))

$$\begin{aligned} (n_1 - n_2) &= 3 \varepsilon_0 E / e r = 4.1 \times 10^4 p, \text{ sphere} \\ &= 2 \varepsilon_0 E / e r = 2.8 \times 10^4 p, \text{ cylinder,} \end{aligned}$$

where $\varepsilon_0 = 1/4\pi$. It will be observed that the ratio E/p is everywhere small, except perhaps when t is very small.

VI. THE DURATION OF A METEOR TRAIL

The scattering of electromagnetic waves has been treated as a volume effect by Lovell and Blackett, and also by Herlofson who considers the effect of plasma resonance in enhancing the scattering (Herlofson 1951).

On the other hand it may be supposed that the scattering occurs, not throughout the expanding cloud of ions, but at that surface where the concentration at that instant has attained the value $n = n_c$, where n_c is the critical electron concentration necessary to reflect electromagnetic waves. As is well known,

when the wave normal is parallel to $\text{grad } n$ the critical concentration is related to the frequency f of the wave by the formula

$$n_c = 1.24 \times 10^4 f^2,$$

where f is expressed in Mc/s. and n_c is the concentration of electrons per cubic centimetre. We therefore adopt a crude simplified picture of the scattering process in which scattering in the expanding column occurs on that cylindrical surface where $n = n_c$. The radius of this surface at first increases to a maximum value and then diminishes to zero. This greatest effective radius of scattering is found immediately from equation (20) when $n_{\max}(r)$ is replaced by n_c . Thus the greatest scattering radius at a frequency f Mc/s. is given by

$$\left. \begin{aligned} r = r_{\max} &= 3 \times 10^{-3} N^{\frac{1}{2}} / f, \\ \text{and the corresponding time } t &= T_1 \text{ is} \\ T_1 &= 2.3 \times 10^{-6} N / (f^2 K) \end{aligned} \right\} \dots\dots\dots (21)$$

After time $t = T_1$ the concentration $n = n_c$ occurs at progressively smaller values of r . The value $n = n_c$ occurs at $r = 0$, according to equation (16) at time

$$t = T_2 = N / (4\pi K n_c) = 6.4 \times 10^{-6} N / (f^2 K) = 2.7 T_1. \dots\dots\dots (22)$$

The duration T of the radio response lies in value between T_1 and T_2 , say $T = 2T_1$.

Scattering from a spherical cloud may be discussed in the same way but in terms of the formulae in Section IV (a).

To obtain a practical estimate of the duration of meteor trails, suppose the scattered response to be received from an ionized column at a height of 85 km. where the pressure of the air is about $p = 5 \times 10^{-3}$ mm. of mercury.

Let $N = 10^{12}$ ions per cm. Put $K = K_1 = 25/p$ on the assumption that the positive ions are ionized air molecules.

Then, from (21) and (22) it follows that if $f = 20$ Mc/s., ($\lambda = 15$ m.)

$$r_{\max} = 150 \text{ cm. ; } T_1 = 1.15 \text{ sec.}$$

If $N = 10^{13}$, $T_{\max} = 470$ cm. ; $T_1 = 11.5$ sec.

These durations are of the order of magnitude commonly observed.

A column formed lower in the atmosphere where $p = 10^{-2}$ mm. with $N = 10^{14}$ would give

$$r_{\max} = 15 \text{ m. ; } T_1 = 300 \text{ sec.}$$

The actual scattering area of a cloud will in fact be much larger than the quantity r_{\max} defined above.

The distance r_{\max} gives the greatest penetration of that portion of the wave front where the wave normal is parallel to $\text{grad } n$. Elsewhere, the wave normal is not in general parallel to $\text{grad } n$ and reflection occurs at distances r greater than r_{\max} . Thus, according to ray optics the scattering cross section should exceed r_{\max} . The behaviour of r_{\max} nevertheless affords a useful criterion for assessing the duration of the group.

It may be concluded therefore that the durations of meteor trails are of the order of magnitude to be expected from the known behaviour of ions and electrons in gases.

VII. REFERENCES

- ABRAHAM, M. (1932).—"The Classical Theory of Electricity and Magnetism." (Revised by R. Becker.) (Blackie & Son Ltd.: London and Glasgow.)
- CARSLAW, H. S., and JAEGER, J. C. (1947).—"The Conduction of Heat in Solids." p. 216. (Clarendon Press: Oxford.)
- COBINE, J. D. (1941).—"Gaseous Conductors." (McGraw Hill Book Co.: New York.)
- HERLOFSEN, N. (1951).—*Ark. Phys. Stockholm* **3**: 15.
- HUXLEY, L. G. H., and ZAAZOU, A. A. (1949).—*Proc. Roy. Soc. A* **196**: 402.
- NIELSEN, R. A., and BRADBURY, N. (1937).—*Phys. Rev.* **51**: 69.
- TOWNSEND, J. S. (1915).—"Electricity in Gases." (Oxford Univ. Press.)
- TOWNSEND, J. S. (1928).—*C.R. Acad. Sci. Paris* **186**: 55.
- TOWNSEND, J. S. (1948).—"Electrons in Gases." (Hutchinson & Co.: London.)

RADIO-FREQUENCY RADIATION FROM THE CONSTELLATION OF CYGNUS

By J. H. PIDDINGTON* and H. C. MINNETT*

[*Manuscript received September 4, 1951*]

Summary

Observations are described of the radiation from portion of the constellation of Cygnus at frequencies of 1210 and 3000 Mc/s. Two sources of radiation were observed at the lower frequency, one being the well-known "radio star", Cygnus-A. The other was a diffuse source of limited extent which might be called a "radio nebula". Neither source could be observed at the higher frequency.

The properties of both sources, particularly their spectra, are discussed and it is shown that earlier discrepancies in observations of the Cygnus region may be explained. The diffuse source coincides in position with the secondary maximum in the lower frequency galactic contours, which Bolton and Westfold (1950*a*, 1950*b*) have interpreted as a spiral arm of the Galaxy. The new evidence suggests that the source is probably due to thermal emission from clouds of ionized interstellar gas, possibly in the region of γ Cygni and having a temperature and electron density of the order of 10^4 °K. and 10 cm.^{-3} respectively.

I. INTRODUCTION

In a recent paper Piddington and Minnett (1951) described measurements of galactic radiation at frequencies of 1210 and 3000 Mc/s. Radiation from the constellation of Cygnus is of particular interest and was therefore omitted from the earlier paper to enable a more detailed discussion to be given in the present communication.

Fluctuations in the intensity of radiation from a small region in Cygnus were first observed by Hey, Parsons, and Phillips (1946) at 64 Mc/s., and were shown by Bolton and Stanley (1948) to originate in a small discrete source. The position of this source was measured by a number of workers whose results differed by more than the limits of error claimed.

A number of equal-intensity contour charts of radiation from the Cygnus region have been published for frequencies between 64 and 480 Mc/s. These show differing features which Bolton and Westfold (1950*b*) have attempted to explain in terms of a secondary maximum of galactic radiation combined with the radiation from the discrete source. They suggested that the secondary maximum was consistent with the Sun being in or near the arm of a spiral galaxy.

The measurements described in the present paper greatly extend the spectrum of the known discrete source in Cygnus, and show that there is a second discrete source which is spread over several square degrees and which

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

had not previously been isolated from the general background of galactic radiation. It is suggested that this may account for the discrepancy in the position measurements of the unresolvable discrete source. Most of the differences between the contour charts at various frequencies are explained by the new evidence, which casts serious doubt on the spiral arm hypothesis. An alternative explanation of the diffuse source may be given in terms of thermal emission from clouds of ionized interstellar gas.

II. EXPERIMENTAL PROCEDURE AND RESULTS

The apparatus and experimental methods used have been described in the earlier paper.

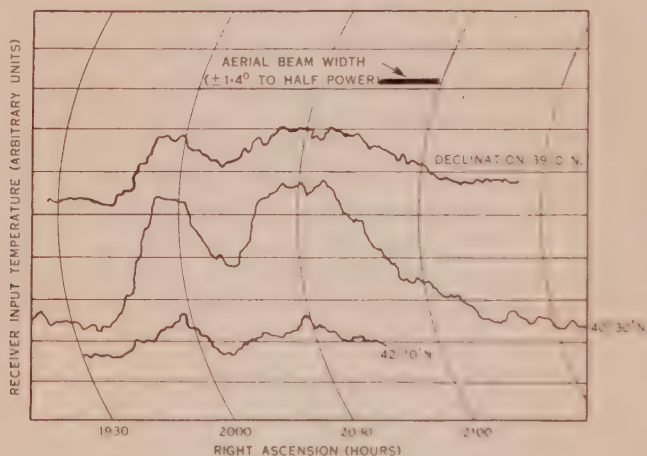


Fig. 1.—Three recordings of the power flux density from Cygnus-A and Cygnus-X (at 1210 Mc/s.).

(a) *The 1210 Mc/s. Measurements*

Measurements at 1210 Mc/s. were made with the 16 by 18 ft. paraboloid described earlier by Piddington and Minnett (1951). The aerial was kept at rest and the portion of the sky being investigated was allowed to drift through the beam, changes in flux density being recorded. The aerial beam width was $\pm 1.4^\circ$ to half-power.

Three typical recordings of the flux density of radiation are shown in Figure 1. They were made with Declinations of the aerial beam of $39^\circ 0' \text{ N.}$, $40^\circ 30' \text{ N.}$, and $42^\circ 10' \text{ N.}$ The width of the aerial beam (appropriate to the given declination) is shown as a thick line above the flux density record. The small fluctuations in the record are of a random nature, being inherent in the equipment which was operating at very high gain. Thus the high sensitivity may be gauged from the fact that the maximum deflection in the three records shown corresponds to a change in the temperature at the receiver input of only about 6° K. In the upper record a drift of the zero level occurred during the observations. Allowance for drifts was made by taking two or more records at the same declination.

It is of interest to compare the traces in Figure 1 with the corresponding traces for the centre of the Galaxy (Piddington and Minnett 1951, Fig. 3). The units of observed intensity in the two figures are the same within about 25 per cent. It is clear that for the frequency and aerial beam width used the intensity from the Cygnus region is comparable with that at the galactic centre.

A series of records of the type shown were made between Declinations 36°N. and 47°N. Excess radiation above the low background level received

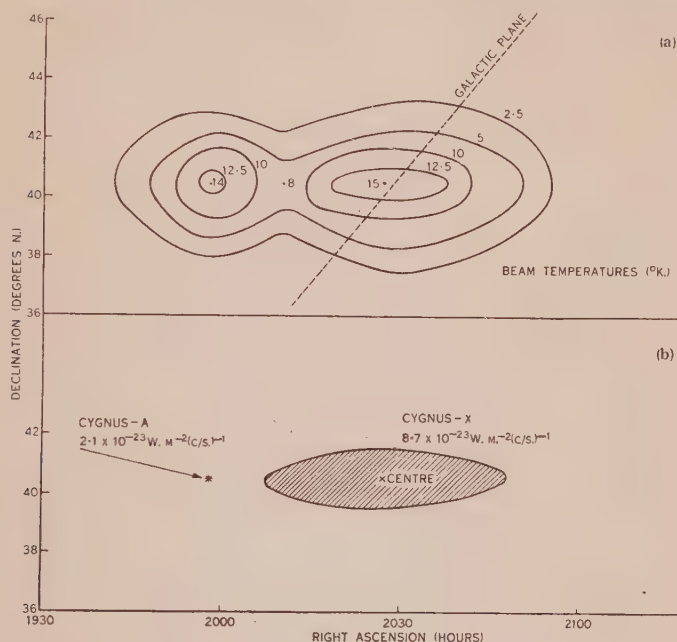


Fig. 2 (a).—Contours of equal aerial beam temperature in the Cygnus region at 1210 Mc/s.

Fig. 2 (b).—Flux densities from the Cygnus sources at 1210 Mc/s.

from regions well away from the galactic plane was only found in the vicinity of Declination 40.5°N. The records have been reduced to a set of equal-intensity contours, shown in Figure 2 (a). The parameter plotted is the "equivalent beam temperature" which is similar to the "equivalent aerial temperature" except that allowance has been made for 35 per cent. losses in side lobes which were directed at cold parts of the sky (see Piddington and Minnett 1951).

Knowing the form of the aerial beam it is possible to reduce the set of contours in Figure 2 (a) to contours showing a distribution of flux density per unit solid angle (that is, to values of brightness) consistent with the experimental results shown in Figure 2 (a) (see for example, Hey, Parsons, and Phillips 1948). The result is shown in Figure 2 (b), where the radiation is seen to originate in two discrete sources, one of which has finite size while the other is too small to be resolved. The position of the small source is the same, within the present

limits of accuracy, as that of the well-known discrete source in Cygnus mentioned in Section I. It is identified with that source and called, after Bolton, Cygnus-A. The second, diffuse source will be referred to here as Cygnus-X and may be the first "radio nebula" to be recognized.

The flux densities from the two sources at 1210 Mc/s. are as follows:

Cygnus-A	2.1×10^{-23} W.m. ⁻² (c/s.) ⁻¹
Cygnus-X	8.7×10^{-23} W.m. ⁻² (c/s.) ⁻¹ .

The limits of error are probably about 25 per cent. in each case.

The position of Cygnus-A is: R.A. 19 hr. 58 min. ± 2 min., Dec. 40.5° N. $\pm 1^\circ$. The position of the centre of Cygnus-X is: R.A. 20 hr. 27 min. ± 3 min., Dec. 40.5° N. $\pm 1^\circ$. The uncertainties quoted are estimates of limits of error based on an examination of a number of records.

(b) The 3000 Mc/s. Measurements

Unfortunately the aerial used for earlier measurements at 3000 Mc/s. could not easily be directed towards the Cygnus region. It was necessary to reduce its size somewhat and hence the overall sensitivity. The sensitivity was further reduced by the fact that the aerial could not be moved during observations and hence the "beam-swinging" techniques described by Piddington and Minnett (1951) could not be used. The sources were allowed to pass through the aerial beam which remained stationary.

The estimated minimum detectable value of flux density from a point source was about 2×10^{-23} W.m.⁻² (c/s.)⁻¹ while the corresponding value for Cygnus-X was about 1×10^{-22} W.m.⁻² (c/s.)⁻¹. The latter value was larger because only about one-third of the total flux density from this extended source was received at any one time and also because the maximum of brightness was broader and took longer to pass through the beam, thus giving less discrimination against random drift.

Observations were made on several occasions but at no time was radiation from either source detected. It was concluded that the upper limits of flux density at 3000 Mc/s. were approximately as follows:

Cygnus-A	2×10^{-23} W.m. ⁻² (c/s.) ⁻¹
Cygnus-X	1×10^{-22} W.m. ⁻² (c/s.) ⁻¹ .

III. THE SMALL DISCRETE SOURCE (CYGNUS-A)

Some of the earlier determinations of the position of Cygnus-A were in error by amounts many times the quoted values of the experimental errors (compare, for example, the results of Bolton and Stanley (1948) with the most recent value quoted below). There is little doubt that large experimental errors were responsible for these discrepancies, and it may be noted that a determination of position from Reber's (1948) contours provided the most accurate result at the time of its publication.

More recent results still fail to show satisfactory agreement. Thus Mills and Thomas (1951) gave the position as R.A. 19 hr. 57 min. 37 sec. ± 6 sec., Dec. $40^\circ 34'$ while Stanley and Slee (1950) give a value of R.A.

19 hr. 58 min. 14 sec., Dec. 40° 36' without quoting probable errors. The values of Right Ascension given differ by 37 seconds or six times the probable error of Mills and Thomas. Other (unpublished) results also show discrepancies. It is suggested that the diffuse source, Cygnus-X, may be responsible for at least some of the variations in the experimentally determined positions of Cygnus-A. Although the former is too large to provide its own observable interferometer pattern, it may cause an apparent displacement in the position of the small source, particularly in the R.A. direction. Such an error would be expected to vary from one interferometer to another.

TABLE 1
RADIATION FROM CYGNUS-A

Observer	Frequency (Mc/s.)	Flux Density (W.m. ⁻² (c/s.) ⁻¹)
Stanley and Slee (1950)	40	3.2×10^{-22}
Stanley and Slee (1950)	60	2.0×10^{-22}
Stanley and Slee (1950)	80	1.4×10^{-22}
Stanley and Slee (1950)	100	1.2×10^{-22}
Stanley and Slee (1950)	160	0.67×10^{-22}
Ryle (1950)	45	1.3×10^{-22}
Ryle (1950)	81	1.4×10^{-22}
Ryle (1950)	214	1.3×10^{-22}
Lovell (1951)	160	0.5×10^{-22}
Mills and Thomas (1951)	97	1.5×10^{-22}
Present paper	1210	0.21×10^{-22}
Present paper	3000	$< 0.2 \times 10^{-22}$

The identification of Cygnus-A with a visible object would constitute a big advance in the field of radio astronomy by showing what type of star or other object was responsible for such powerful radio emission. Unfortunately, as Mills and Thomas (1951) have shown, the present positional uncertainties are far too great to allow even tentative identification. In the case of an isolated source these uncertainties may be reduced by the use of improved experimental technique. However, in the case of Cygnus-A the proximity of the diffuse source Cygnus-X may constitute a limit to the accuracy obtainable. An estimate of the effect of Cygnus-X on a given interferometer system can be made if enough data are available, thereby reducing the limits of uncertainty.

The flux density of radiation from Cygnus-A has been measured by several observers at frequencies between 40 and 214 Mc/s. with the results listed in Table 1.

These values, multiplied by a factor of 10 to avoid overlapping the other curves, are plotted in Figure 3. Limits of error of Stanley and Slee and the present authors are shown by thick vertical lines. Where no limits are given the results are shown by dots and the initials of the observer. A curve has been

drawn to fit the data as well as possible. With two exceptions the results fall on or close to this curve. However, Ryle's results at 45 and 214 Mc/s. are displaced from the curve by factors of 2.1 and 2.8 respectively.

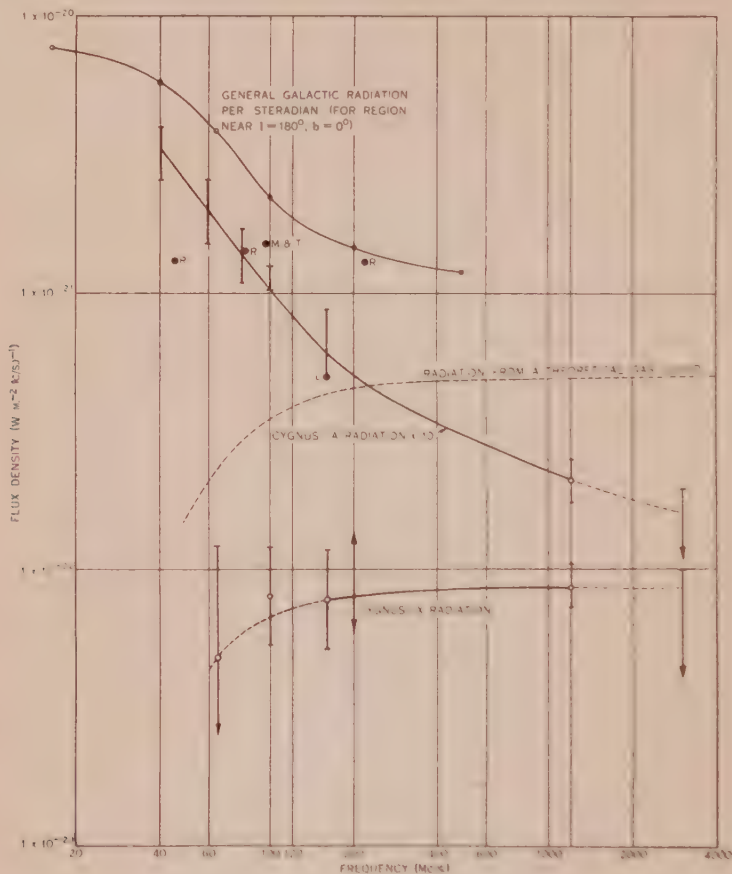


Fig. 3.—Radio-frequency emission spectra from the Milky Way (near $l=180^\circ$), from the sources Cygnus-A and Cygnus-X and from a theoretical gas cloud. Values of flux densities for Cygnus-A are multiplied by 10 to separate the curves.

IV. THE DIFFUSE SOURCE (CYGNUS-X)

The measured value of the total flux density of radiation from Cygnus-X at a frequency of 1210 Mc/s. and an upper limit at 3000 Mc/s. were given in Section II. As shown below, the flux densities at other frequencies may be estimated approximately from published data, so that an emission spectrum between 64 and 3000 Mc/s. may be drawn. The size and shape of the source are also considered in greater detail.

(a) The Spectrum Curve

Measurements of galactic radiation from the Cygnus region have been made by Hey, Parsons, and Phillips (1948) at 64 Mc/s., by Bolton and Westfold

(1950*a*) at 100 Mc/s., by Reber (1944) and Lovell (1951) at 160 Mc/s., by Allen and Gum (1950) at 200 Mc/s., and by Reber (1948) at 480 Mc/s. The results, together with those of the present paper, are examined in Appendix I in an attempt to determine the spectrum curve of the total flux density of radiation from Cygnus-X.

The problem of determining the total emission from a region of given extent may be divided into three cases depending on the relative extents of the source and the aerial beam used. The cases where the source is very large or very small compared to the aerial beam are relatively simple to solve. Unfortunately the data concerning Cygnus-X are mainly examples of the third case where the dimensions of the source and aerial beam are of the same order. Interpretation of some of the data was difficult and limits of accuracy could only be estimated roughly. No useful estimate of total emission could be made from the 480 Mc/s. data. The estimated values of flux densities are shown as circles in Figure 3 and the limits of uncertainty are indicated by the thick vertical lines. Where the latter are terminated by an arrow no reasonable estimate of uncertainty could be made. A spectrum curve fitting the points as well as possible is shown in the figure. At the lower frequencies a decrease of intensity is shown (dotted portion of curve). The calculations are not sufficiently accurate to determine conclusively whether this occurs, but it is interesting to examine the consequences of such a decrease in connection with a thermal theory of origin. This is discussed in Section V (*b*).

For purposes of comparison, a spectrum curve of general galactic radiation is also shown in Figure 3. This applies to a region in the plane of the Galaxy away from the galactic centre ($l=180^\circ$), the intensity being in units of flux density per steradian. It is considered to be typical of galactic radiation (Piddington, in press).

(*b*) *The Size and Shape of the Source*

As indicated in Section II the experimental results summarized graphically in Figure 2 (*a*) may be explained in terms of the two sources shown in Figure 2 (*b*). The brightness of the Cygnus-X source is arbitrarily assumed to be uniform over the area shown. Although Figure 2 (*b*) *would* explain the observations, it is unlikely that the real source has the shape and uniform brightness shown. Resolution beyond the limits shown is impossible with the aerial beam used and it is likely that within the shaded area indicated the brightness varies considerably. However, the following features of the source appear to be established.

(i) The direction of maximum aerial beam temperature is R.A. 20 hr. 27 min., Dec. 40.5° N. with an uncertainty in position of less than 1° . In galactic coordinates the position is $l=48^\circ$, $b=-1^\circ$, almost in the plane of the Galaxy.

(ii) The source extends for some 10 degrees in the direction of changing right ascension. The uncertainty in this estimate is difficult to assess but is probably not more than two or three degrees.

(iii) The source extends only over a small angle in the direction of changing declination. This result depends on the assumption that Cygnus-A is a point source. The relative rates at which flux density from the two sources changed

with declination were compared. The Cygnus-X flux seemed to fall off only slightly more slowly than the Cygnus-A flux as the declination was changed from 40.5°N . This is clearly illustrated by the records of Figure 1. It was concluded that the extent of Cygnus-X along a line of constant right ascension was probably between one and two degrees.

(iv) Cygnus-X is not composed of a small number (say three or four) of very small sources. The evidence for this is in the 100 Mc/s. interferometer records of Bolton and Stanley (1948), Mills and Thomas (1951), and others. Not only would a few small sources have produced an interference pattern additional to that of Cygnus-A, but they would have prevented even that degree of agreement which has so far been achieved between the various determinations of the position of Cygnus-A.

(c) An Explanation of Earlier Experimental Results

Early observations of the Cygnus region showed only one area of maximum intensity at 64 Mc/s. (Hey, Parsons, and Phillips 1948) and 160 Mc/s. (Reber 1944), while a later plot of Reber (1948) at 460 Mc/s. showed two. Interferometry records indicated the presence of only one small (in size) discrete source in the region (Bolton and Stanley 1948). Reber (1948) suggested that the explanation of these apparently conflicting results lay either in a relative decrease of emission from one source at the lower frequencies or in the fact that one source was diffuse ("perhaps a degree in diameter"). Bolton and Westfold (1950*b*) adopted the second of these possibilities and attempted to explain the results in terms of two sources, the well-established Cygnus-A source and a diffuse source having much the same emission spectrum as the general galactic radiation. It is now clear that both of Reber's suggestions are necessary to explain the various experimental results and that, when allowance is made for different source sizes and spectra as well as aerial beam widths, most of the data can be reconciled.

There is, however, a remaining discrepancy: the plots of observed flux density of Reber (1948) at 480 Mc/s. and the present authors (see Fig. 2 (*a*)) at 1210 Mc/s. show a different orientation of the Cygnus-X source. The directions of elongation of the sources are perpendicular to one another. The beam widths used were about the same but there is a $2\frac{1}{2}$ to 1 difference in frequency. No explanation of the discrepancy is suggested, but further observations seem desirable.

V. THE ORIGIN OF THE CYGNUS-X RADIATION

In the light of the above results on the spectrum and shape of the diffuse source in Cygnus, it is of interest to speculate on the mechanism, and, as far as possible, on the actual object responsible for the emission.

(a) Origin in a Spiral Arm of the Galaxy

Bolton and Westfold (1950*b*) have suggested that the source may be identified with a spiral arm of the Galaxy. They considered that evidence in favour of such an origin was provided by the various sets of contours of Hey, Parsons,

and Phillips, Reber, and Bolton and Westfold. After eliminating the effect of Cygnus-A, these show evidence of a diffuse maximum of radiation in the Cygnus region. Bolton and Westfold consider that this distribution can be interpreted as locating the Sun in or near the arm of a spiral galaxy. However, in analysing the observations they do not appear to have taken full account of aerial beam widths and other factors discussed in Appendix I. For example, they concluded that at 64 Mc/s. the flux density of radiation from the diffuse source was about four times that from Cygnus-A. In Appendix I it is shown that the flux density from the diffuse source is unlikely to be as great as that from Cygnus-A and may well be much less.

The strongest evidence against the hypothesis that the diffuse source in Cygnus is a spiral arm (or any other substantially large portion of the Galaxy) is given by a comparison of its spectrum with that of general galactic radiation (Fig. 3). The latter spectrum changes steadily and by a limited amount in passing from the galactic centre to the poles (Piddington, in press). In all cases it differs from the spectrum of Cygnus-X to an extent which is too great to be due to experimental errors. It seems unlikely that a large section of galactic material, such as a spiral arm, would have an averaged spectrum of emission differing so completely from that of the remainder of the Galaxy.

Moreover, the evidence at both 1210 and 480 Mc/s. shows that the Cygnus-X source is inclined to the galactic plane. Thus Figure 2 indicates that the source at 1210 Mc/s. is long and narrow with its major axis tilted at about 50 degrees to the galactic, whereas a spiral arm might be expected to have its greatest extent in that plane. Actually however, the size and intensity of the source are rather insignificant compared with the Galaxy as a whole. For example, at 100 Mc/s. the total flux density from the source is about $7 \times 10^{-23} \text{ W.m.}^{-2} (\text{c/s.})^{-1}$ and about $3 \times 10^{-20} \text{ W.m.}^{-2} (\text{c/s.})^{-1}$ from the whole Galaxy. Thus the source represents less than $\frac{1}{4}$ per cent. of the total and even if it were a projecting portion of the Galaxy, it would hardly merit the title "spiral arm".

The evidence given above seriously weakens the spiral arm hypothesis. In the following section it is suggested that the radiation may have a thermal origin in clouds of ionized gas. This provides a possible alternative explanation, although the experimental evidence at present is insufficient to draw definite conclusions.

(b) A Thermal Origin in a Cloud of Ionized Gas

It has been shown by Piddington (in press) that the observed galactic radiation in the radio spectrum may be explained in terms of radiation by stars or other small sources and both emission and absorption by relatively small (on a galactic scale) clouds of ionized, interstellar gas. The temperatures and densities required are about $1 \times 10^4 \text{ }^\circ\text{K.}$ and 20 cm.^{-3} respectively, in satisfactory agreement with corresponding estimates from visual data. It was also shown that there was a reasonable chance that one or more of these clouds might be observed if a suitable radio telescope were employed.

The emission spectrum of a gas cloud was discussed and shown to exhibit uniform emission over the (high frequency) range where the optical depth was

small together with a sharp decrease in the region of unit optical depth and negligible emission at lower frequencies. The theoretical spectrum of a cloud of uniform thickness is shown in Figure 3. The optical depth of the cloud at any frequency is assumed to be equal over the whole of its surface and the flux density and optical depth at a given frequency are chosen arbitrarily.

It is suggested that one or more of these clouds of ionized gas might be the source of the Cygnus-X radiation. Approximate estimates of the average electron temperature and density in the clouds may be made from the radio results if certain assumptions are made. The average brightness temperature of the diffuse source is given by

$$T_b = \frac{S\lambda^2}{2k\Omega},$$

where S is the observed power flux density, λ the wavelength, k the Boltzmann's constant, and Ω the solid angle subtended. The solid angle of the distribution shown in Figure 2 is about 5×10^{-3} steradians. Thus for a frequency of 1210 Mc/s. we find $T_b = 40^\circ \text{K}$.

It will be tentatively assumed that the Cygnus-X spectrum is as shown in Figure 3 although, as stated in Section IV (a), the accuracy of the calculations is insufficient to determine whether there is a decrease of intensity at the low-frequency end of the curve or not. If it is subsequently shown that the decrease starts at a different frequency to that indicated or else not at all, then the following quantitative results must be modified or abandoned.

The spectrum curve shown is consistent with thermal radiation from ionized gas having an "average" optical depth of about 1.2 at 64 Mc/s.* and of 3.4×10^{-3} at 1210 Mc/s. Since the electron temperature T_e is related to the brightness temperature T_b and the optical depth τ by

$$T_b = T_e(1 - e^{-\tau}),$$

we find $T_e = 1.2 \times 10^4^\circ \text{K}$. Although this estimate is necessarily very rough, it is in good agreement with the earlier value based on other radio data (Piddington, in preparation) and with the value estimated from visual observations of H II atmospheres (Strömgren 1918). It may be noted that if the intensity starts to decrease at a higher frequency than that shown in Figure 3, a lower value of T_e would result. ($T_e \sim 3000^\circ \text{K}$. if the value of flux density has fallen to half at 100 Mc/s.)

To determine the average electron density in the emitting gas clouds an estimate of the thickness of the clouds is necessary. In the following section a tentative identification of the radiating object is made with the gas clouds illuminated by γ Cygni. This star is known, from parallax determinations, to lie at a distance of about 120 parsecs. Assuming that the surrounding gas clouds lie at about the same distance, the size of the radiating object would be

* It is assumed here that the optical depth is inversely proportional to the square of the radio frequency. This may not be quite true (Piddington, in press) but is sufficiently accurate for present purposes.

about 4 parsecs by 20 parsecs and its average thickness might be expected to be of the same order, about 10 parsecs. The optical depth is thus

$$\tau = 3 \times 10^{19} \kappa,$$

where κ is the absorption coefficient (c.g.s. units) and is given by (Piddington, in press)

$$\kappa = \frac{0.14 N^2}{f^2 T_e^{3/2}},$$

where N is the (average) electron density and f the wave frequency. Using the 1210 Mc/s. data we find a value of $N = 40 \text{ cm.}^{-3}$.

If the identification of the source with the γ Cygni clouds must be rejected, then the best estimate of distance of the object is probably about 10^4 parsecs. This is based on the fact that the source, like a number of others (Piddington and Minnett 1951), lies very close to the galactic plane and is likely, therefore, to be situated at a distance large compared to the thickness of the Galaxy. Adopting a distance of 10^4 parsecs (which is approximately an upper limit also), the corresponding value of N is 4 cm.^{-3} . Either of these estimates of N is in reasonable agreement with the previous values derived from optical (Strömgren 1948) or radio data.

(c) *A Possible Identification of the Source*

A tentative identification of the diffuse Cygnus source has been made with the bright galactic nebulae surrounding γ Cygni. The identification is based on the fact that a number of bright nebulae are distributed in the region occupied by the source and these, like the source, appear to be most extensive along a line of fixed declination in the region of 40°N . Beyond this slight similarity in shape, the general outline bears no similarity to the simplified source shape shown in Figure 2 (b).

The nebulae concerned are listed by Cederblad (1946) as (his) number 176 and include IC 1318 and NGC 6914. They have been discussed by Hubble (1922) who suggested that they were illuminated by γ Cygni. It should be noted that this star, type F8p, would be quite incapable, according to Strömgren's theory, of maintaining any widespread H II atmosphere such as the radio data indicate. Other stars of types O and B obscured by the gas clouds may, however, be responsible.

It was pointed out by Piddington and Minnett (1951) that the proximity of a number of discrete sources to the galactic plane was evidence that they were probably at great distances from the Earth. If the above identification of Cygnus-X is verified, then in this case the proximity to the galactic plane is coincidental, the objects being relatively close to the Earth.

VI. ACKNOWLEDGMENTS

The authors are grateful to Mr. J. V. Hindman for assistance with some of the observations and to Mr. W. N. Christiansen for the loan of his 18 by 16 ft. aerial. They are also indebted to Mr. B. Y. Mills and Dr. J. I. Pawsey for helpful criticism in the preparation of this paper.

VII. REFERENCES

- ALLEN, C. W., and GUM, C. S. (1950).—*Aust. J. Sci. Res. A* **3**: 224.
 BOLTON, J. G., and STANLEY, G. J. (1948).—*Aust. J. Sci. Res. A* **1**: 58.
 BOLTON, J. G., and WESTFOLD, K. C. (1950a).—*Aust. J. Sci. Res. A* **3**: 19.
 BOLTON, J. G., and WESTFOLD, K. C. (1950b).—*Aust. J. Sci. Res. A* **3**: 251.
 CEDERBLAD, S. (1946).—Studies of bright diffuse galactic nebulae. *Medd. Lunds Abstr. Obs. Ser. 2*, No. 119.
 HEY, J. S., PARSONS, S. J., and PHILLIPS, J. W. (1946).—*Nature* **158**: 234.
 HEY, J. S., PARSONS, S. J., and PHILLIPS, J. W. (1948).—*Proc. Roy. Soc. A* **192**: 425.
 HUBBLE, E. (1922).—*Astrophys. J.* **56**: 400.
 LOVELL, A. C. B. (1951).—*Discovery* **12**: 7.
 MILLS, B. Y., and THOMAS, A. B. (1951).—*Aust. J. Sci. Res. A* **4**: 158.
 PIDDINGTON, J. H.—The origin of galactic radio-frequency radiation. *Mon. Not. R. Astr. Soc.* (in press).
 PIDDINGTON, J. H.—A galactic model of the sources of radio-frequency radiation (in preparation).
 PIDDINGTON, J. H., and MINNETT, H. C. (1951).—*Aust. J. Sci. Res. A* **4**: 459.
 REBER, G. (1944).—*Astrophys. J.* **100**: 279.
 REBER, G. (1948).—*Proc. Instn. Radio Engrs. N.Y.* **36**: 1215.
 RYLE, M. (1950).—*Rep. Phys. Soc. Progr. Phys.* **13**: 184.
 STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res. A* **3**: 234.
 STRÖMGREN, B. (1948).—*Astrophys. J.* **108**: 242.

APPENDIX I

THE SPECTRUM OF RADIATION FROM CYGNUS-X

(a) *The 64 Mc/s. Data*

Hey, Parsons, and Phillips (1948) have published a set of equal brightness contours of galactic radiation at 64 Mc/s. The relevant portion is reproduced in Figure 4 (a); the maximum shown presumably corresponds to the excess radiation (above the general background) from the sources Cygnus-A and Cygnus-X. This excess was determined by first redrawing the contours with the maximum "shaved off", as well as could be estimated, from the background. The result is shown in Figure 4 (b). The difference in total flux densities for the two figures was then found, by a process of graphical integration, to be 2.3×10^{-22} W.m.⁻² (c.s.)⁻¹, which may be compared to a flux density from Cygnus-A at the same frequency of 1.8×10^{-22} W.m.⁻² (c.s.)⁻¹ found from the curve in Figure 3. The flux density from Cygnus-X is therefore about 0.5×10^{-22} W.m.⁻² (c.s.)⁻¹.

The method is not very accurate. However, if the uncertainty in the graphical determination of the flux density from both sources (which has been repeated several times) is about ± 20 per cent. and the accuracy of the Cygnus-A value is ± 10 per cent., then the value of Cygnus-X should be between zero and an upper limit of 1.2×10^{-22} W.m.⁻² (c.s.)⁻¹.* One reason for believing that it is finite is that the peak shown by Hey, Parsons, and Phillips has its maximum

* It should be noted that the flux density from Cygnus-A is taken from the curve in Figure 3, which gives the best fit for all the experimental results. If Ryle's result at 45 Mc/s. were given full weight, the result would have to be increased to 1.0×10^{-22} W.m.⁻² (c.s.)⁻¹ with an upper limit somewhat higher. The weight of evidence seems to be in favour of the value derived from the smooth curve.

between the positions of Cygnus-A and Cygnus-X. The reason why Hey, Parsons, and Phillips have shown a single maximum where there are two is the difficulty of interpreting fine structure when using a wide (about $\pm 10^\circ$) aerial beam.

Bolton and Westfold (1950*b*) concluded that the flux density from the diffuse maximum (our Cygnus-X) was a much larger proportion of the total radiation from the region concerned than that found above. Bolton (personal communication) has informed the authors that this was done because he believed that the variable part of the Cygnus-A radiation might not have been allowed for in the contours of Hey, Parsons, and Phillips.

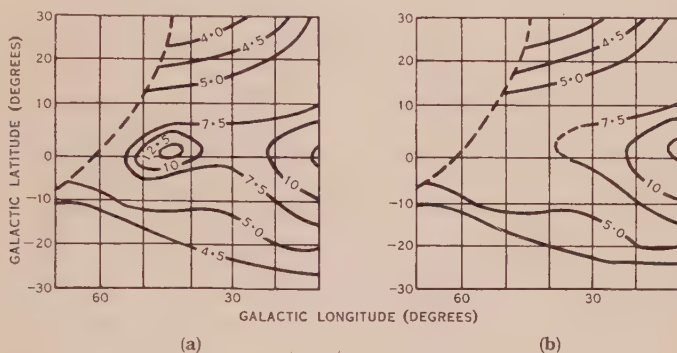


Fig. 4.—Equal brightness contours of galactic radiation at 64 Mc/s.
 (a) As given by Hey, Parsons, and Phillips (1948).
 (b) After modification by removing the Cygnus sources.

(b) The 100 Mc/s. Data

It is not possible to estimate the excess radiation from the Cygnus region from Bolton and Westfold's (1950*a*) contours, which are distorted because of the proximity of the aerial beam to the horizon. However, they have made a series of measurements of the Milky Way in Cygnus rising over the sea; one of these is shown in their Figure 4. The others (unpublished but kindly lent to the present authors) are similar except that the interference pattern due to Cygnus-A is smaller or absent altogether. It should be noted that the main part of the maximum of intensity shown is due to the beam crossing the Milky Way, and would be obtained for any galactic longitude. The "local diffuse maximum" is a fraction only of this value.

The fluctuation of aerial temperature due to the Cygnus-A source is 600°K. (peak to peak) corresponding to a free space flux density of $1.25 \times 10^{-22} \text{ W.m.}^{-2} (\text{c/s.})^{-1}$.* After eliminating the effect of Cygnus-A, the maximum aerial temperature is about 200°K. higher than the average value for two positions on each side (where no radiation from the Cygnus region is received). Hence the flux density of Cygnus-X is about $0.8 \times 10^{-22} \text{ W.m.}^{-2} (\text{c/s.})^{-1}$

* The power received from a point source in the presence of a perfectly reflecting sea varies between zero and four times the free space value. A diffuse source produces small or zero fluctuations in received power with a mean value of twice that for free space.

with an uncertainty which is estimated to be about ± 50 per cent. No appreciable correction is necessary to allow for the fact that the radiation from Cygnus-X is spread over about $\pm 5^\circ$ of the $\pm 8.5^\circ$ aerial beam used.

(c) *The 160 Mc/s. Data*

Reber's (1944) 160 Mc/s. results are in the form of contours of equal intensity of radiation, averaged over the solid angle of his aerial beam. The radiation from the Cygnus sources may be found, as shown by Bolton and Westfold, by comparing the observed maximum in Cygnus with that in Cassiopeia. It is first necessary, however, to decide what allowance to make for the decreased effective aerial gain due to the spread of the pair of Cygnus sources.

The aerial beam width may be estimated from published dimensions and wavelength as at least $\pm 6^\circ$. Assuming that the size and positions of the sources are as shown in Figure 2 and that the flux densities from the two sources are equal, then the loss due to spreading of the sources is about 20 per cent. A wider aerial beam or unequal flux densities will reduce this loss.

Of the total observed radiation from the Cygnus maximum shown by Reber, and Bolton and Westfold (1950*b*) have shown that about 1.4 (Reber's) units are due to Cygnus-A and the remainder to a second source (our Cygnus-X). The total radiation is 3.4 units less 0.7 units for the more or less uniform background; the remaining 2.7 units are increased to 3.3 units to allow for the narrow aerial beam. Thus the flux density from Cygnus-X is about $3.3 \div 1.4 = 1.9$ units. As the flux density from Cygnus-A is 6×10^{-23} W.m.⁻² (c.s.)⁻¹ (see Fig. 3), that from Cygnus-X will be about 8×10^{-23} W.m.⁻² (c.s.)⁻¹ or perhaps somewhat less if too much allowance was made for beam width effects. The limits of error are again taken as 50 per cent.

A single observation of the Cygnus sources at 160 Mc/s. has recently been published by Lovell (1951) which may be compared with the 1210 Mc/s. observations illustrated in Figure 1. It is difficult at the lower frequency to separate the diffuse source from the general background radiation which is clearly in evidence in Lovell's curve. However, if Cygnus-X is assumed to have the dimensions given in Figure 2 (*b*), it is possible to estimate its flux density. Allowing for the decreased beam width and increased Cygnus-A radiation, there appears to be some diminution in flux density from Cygnus-X between 1200 and 160 Mc/s. in general agreement with the results illustrated in Figure 3.

Lovell's curve may also indicate a tendency for the diffuse source to split when examined by a narrower aerial beam.

(d) *The 200 Mc/s. Data*

Allen and Gum (1950) have plotted the distribution of received power (their Fig. 5), using a beam of $\pm 12.5^\circ$. They find a maximum in Cygnus of about 2.1 units. Their aerial gain may be estimated as 35 (times an isotropic radiator) and their base level as about 1.7 units, so that the total flux density from the Cygnus sources is

$$\frac{(2.1 - 1.7) \times 10^{-21} \times 4\pi}{35} = 14.4 \times 10^{-23} \text{ W.m.}^{-2} \text{ (c.s.)}^{-1}.$$

Allowing a contribution of 5×10^{-23} W.m.⁻² (c/s.)⁻¹ from Cygnus-A (Fig. 3), the flux density from Cygnus-X is about 9×10^{-23} W.m.⁻² (c/s.)⁻¹. The estimates, both of the base level of Allen and Gum's radiation and of the aerial gain, are open to considerable doubt; no estimate is made of the limits of error.

(e) The 480 Mc/s. Data

Reber's (1948) contours in the region concerned are incomplete, extending only to about Dec. 44°N. Also they appear to be inconsistent, both with the present results (Section IV (c)) and between themselves (compare Reber's Figs. 3 and 5). For these reasons it is not feasible to estimate the flux density of Cygnus-X from his data.

(f) The 1210 Mc/s. Data

It has been shown in Section II (a) that the total flux density of radiation from Cygnus-X at 1210 Mc/s. is 8.7×10^{-23} W.m.⁻² (c/s.)⁻¹ with an estimated uncertainty of about 25 per cent.

THE POSITION AND MOVEMENT ON THE SOLAR DISK OF SOURCES OF RADIATION AT A FREQUENCY OF 97 Mc/s.

III. OUTBURSTS

By RUBY PAYNE-SCOTT* and A. G. LITTLE*

[*Manuscript received July 20, 1951*]

Summary

The apparent position and polarization of the source of solar radiation at 97 Mc/s. have been measured during six outbursts, for five of which data on the accompanying solar flares are available. Initially the apparent position of outburst and flare almost coincide, the outbursts usually being rather closer to the solar limb than the flare. As the outburst proceeds, its position rapidly changes, the apparent movement usually being towards, and sometimes off, the solar limb. The initial apparent displacement towards the solar limb and the subsequent movement can be explained if we assume that the outburst radiation is excited by some physical agency originating at the same time, and in the same region, as the flare, and moving outward through the corona. The velocity of such an agency would need to be between 500 and 3000 km./sec., and it is suggested that the corpuscular streams associated with sudden commencement terrestrial magnetic storms are the initiators of outbursts.

The polarization of the outbursts in the early stages is random. Often a second increase occurs, with elliptical, usually circular, polarization. On two occasions linear polarization has been observed in this later stage of an outburst. The relation between the changes in polarization and the associated changes in apparent position conform to the rule given in Part II of this series, that right-hand circular polarization arises in the region above a south magnetic pole and left-hand above a north pole, while linear polarization appears to originate above the central region of a bipolar group, the E -vector having a direction approximately along the axis of the spot-group.

I. INTRODUCTION

Occasionally huge sudden increases occur in the radiation from the Sun observed at radio frequencies, lasting for some minutes and rising on occasions to intensities of at least a million times that of the thermal radiation observable from the "quiet" Sun. It was early observed (Hey 1946) that such outbursts often accompanied solar flares, and one of the present authors has shown (Payne-Scott 1949) that there is a very good correlation between large outbursts on 85 and 60 Mc/s. and major radio fade-outs, although Allen (1947) found only slight correlation between outbursts at 200 Mc/s. and all observed solar flares (both large and small). Wild (1950), recording over the frequency range 70–130 Mc/s., has observed during outbursts a drift of the spectral features with time towards the lower frequencies, and interprets this drift as due to the movement of a physical agency which initiates the outburst during its passage through the solar atmosphere.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

In view of the observed correlation between flares and outbursts, and the suggestion that some agent moving out through the solar atmosphere is involved, it is of particular interest to have continuous measurements of the position of the radiating source during outbursts. Such measurements can be made with the interferometer, described in Part I of this series (Little and Payne-Scott 1951), which is particularly designed for the study of short-lived disturbances, enabling a complete measurement of position and polarization of the radiating source to be made in about a second. The present paper describes such studies of the changes in position and polarization of the source during outbursts, and discusses their significance.*

II. EQUIPMENT

The interferometer described in detail in Part I of this series (Little and Payne-Scott 1951) enables the apparent position of a radiating source at 97 Mc/s. to be determined on a line parallel to the hour-circle through the centre of the solar disk, with an accuracy of about ± 2 minutes of arc. The polarization is also determined as circular, elliptical, linear, random, or a combination of these.† A single measurement of position and polarization can be made in a second. Plate 1 shows the type of record from which measurements are made. The first two photographs in each set are taken with the spaced aerials parallel and then mutually at right angles, and the last with aerials again parallel but with a different spacing. The set shown was taken during the outburst of February 17, 1950, and the gradual appearance of a sine wave in the central photo of each set, displaced in phase by a quarter-wavelength from that in the first photo, shows the gradual change in polarization from random to circular. The deviations from a straight line in the central photo of the first set are due to irregularities in the phase changer.‡ In addition, a continuous record of intensity is made on an independent aerial system, adapted to receive right-handed and left-handed circularly-polarized radiation in alternate half-minutes, thus providing a check on the measurements of polarization, as well as the intensity. This system operates at 98 Mc/s.

III. EXPERIMENTAL RESULTS

Observations have been made on the interferometer during six large sudden increases of solar noise. Of these, five (September 5, 1949, February 17 and 22, 1950, August 5 and 11, 1950) are known to have been accompanied by optical flares and radio fade-outs. The remaining increase (June 29) was of shorter duration. We have no optical information on this occasion, but a radio fade-out

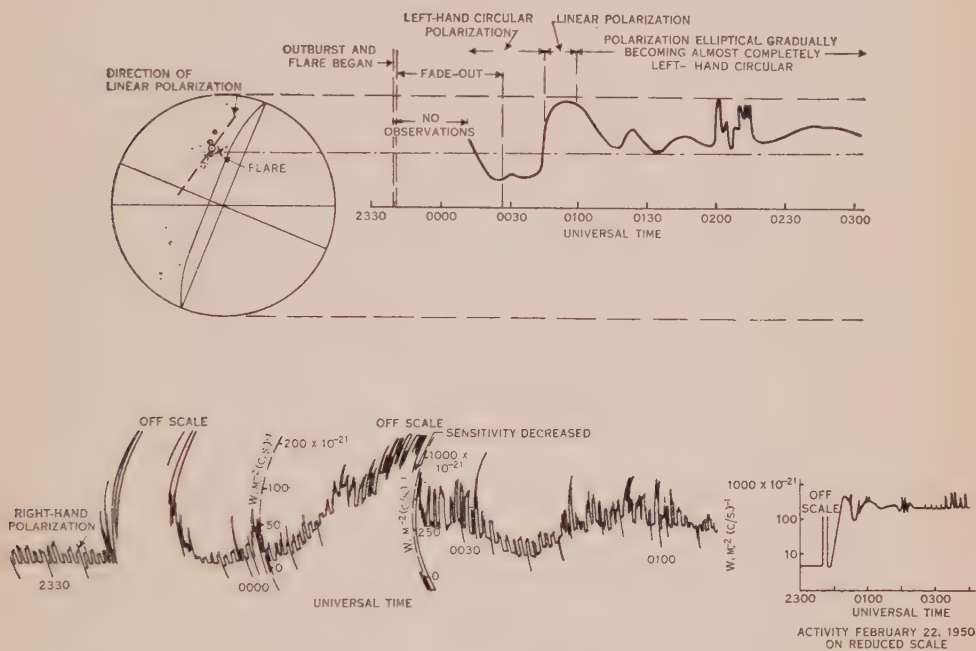
* Some of this work has already been described at the meeting of U.R.S.I. at Zürich in September 1950.

† The ordinary radio convention is used for the direction of rotation in elliptical polarization, the polarization being right-handed when the direction of rotation of the vector in a fixed plane normal to the direction of propagation appears clockwise to an observer looking along the direction of propagation of the ray.

‡ The phase changer in use at this time was replaced a little later by one having a greater length and fewer irregularities.

was observed. The plots of the observed positions of the sources of 97 Mc/s. radiation against time, with a note of the polarizations, are shown in Figure 1, together with sketches of the visible solar disk. For each outburst the position of the corresponding flare is marked with a cross.

This figure also shows the record of intensity on 98 Mc/s. for each outburst. The downward kicks on the trace are time marks, put on automatically every 5 minutes, with three marks at minute intervals denoting the hour. The aerials are switched so as to receive right- and left-handed circular polarization in



FEBRUARY 22, 1950

Fig. 1 (c)

alternate half-minutes, the left-handed component being received in the half-minute after each time mark. The total incident power is the sum of that received on the aerials in each of the two positions, and hence for random polarization is twice that received in either position.

Table 1 shows the available optical, radio, and magnetic data, and some of this is repeated in Figures 1 (a) to 1 (f). In the upper half of each figure is shown the visible solar disk, with plot of measured distance from hour-circle through centre of Sun against time of source of 97 Mc/s. radiation ; in the lower half the intensity record on 98 Mc/s. is set out.

(a) The Relation between the Times of Commencement of Flare, Radio Fade-out, and Outburst

The commencing times of the outbursts were, except on one occasion, within a few minutes of these associated fade-outs and flares, the outburst occurring

TABLE I
OUTBURSTS OF SOLAR RADIO NOISE AND ASSOCIATED FLARE, FADE-OUT, AND MAGNETIC STORM DATA

Date	Outburst on 98 Mc/s.		U.T.	Flare Data*						Radio Fade-outs†		S.C. Magnetic Storms reported by Watheroo Observatory, W.A.			
	Began	Ended		Observatory	Began	Max.	End	Importance	Position	Remarks	Duration	f_{min} (Mc/s.)	Began	Ended	Max. K Index
Sept. 5, 1949	0214	0233†		Kodaikanal, India	0217	0255	2	178., 7E.			0210-0250	4-5			
Feb. 17, 1950	0127	0202		Carter, N.Z. Kodaikanal, India	0130 0210	0305	3	10N., 37E.			01254-0235	Complete fade			
Feb. 21-22, 1950	2340	No definite end; high activity continued at least 24 hours		Carter, N.Z.	2137 2226 2358 0110	2141 2306 2343 0110	1+ 1- 3 2	8N., 25W. 13N., 28W. 13N., 33W. 10N., 24W.	Could perhaps be a 3+ flare. Continual brightening and fading of hydrogen in this group and groups 1426 and 1436 all day	2341-0027	Complete to 4.5	1041 Feb. 23	1200 Feb. 24	6	
				Commonwealth Observatory, U.S.A.	0115 0257 0000	0200 0331 0040	1 2+ 2	10N., 16W. 13N., 29W. 12N., 28W.							
June 29, 1950	0135	0137‡		Mt. Wilson, U.S.A.	2340	2346	3+	11N., 26W.	In two distinct portions; stronger at about 31W., weaker at about 20W.	0134-0207	2-1				
Aug. 3, 1950	2332	0007		Commonwealth Observatory, Australia	0100		1	78., 5W.		2332-0105	5-5	1055 Aug. 7	1200 Aug. 8	7	
Aug. 11, 1950	00304	0107		Mt. Wilson, U.S.A.	0100	0111	2	14N., 74E.		0050 0110 10	2-1				

* Times in italics are actual times of beginning and ending.

† Duration of fade-out taken from our own records. ‡ Radio data provided by the Commonwealth Observatory. Time of start correct to 1 min. (except on August 11, 1950) time of ending correct to 1-30 min.

a minute or two after, or occasionally at the same time as, the fade-out or flare. The exception occurred on August 11 (Fig. 1 (f)) when there was a delay of $16\frac{1}{2}$ minutes between the onset of the flare and of the outburst. The time of onset of the fade-out is not certain, but it was certainly also considerably later than that of the flare. This outburst, as we shall see later, showed a direction of apparent movement of the radiating source opposite to that in the others.

(b) The Initial Source Position and Polarization of Outbursts Observed from Commencement

In four cases (Figs. 1 (a), 1 (b), 1 (d), and 1 (f)) position and polarization measurements were made at or near the beginning of the outburst. In each of these the behaviour was similar. The polarization was initially random and the initial position on the solar disk of flare and outburst almost coincided, the outburst being slightly closer to the solar limb than the flare. When the outburst was associated with a flare or spot group near the limb (June 29 and August 11; Figs. 1 (d) and 1 (f)) this displacement was more marked.

(c) Subsequent Source Movement and Polarization of These Outbursts

As the outburst progressed, its observed position changed rapidly, the apparent movement in three cases being towards the solar limb. The exceptional case (August 11; Fig. 1 (f)), in which the apparent movement was towards the central meridian, was the outburst already mentioned which occurred much later than the flare presumed to be associated with it. The polarization always remained random for some time after the beginning of an outburst, and on two occasions (June 29 and August 11; Figs. 1 (d) and 1 (f)) no other polarization was detectable. In the other two cases (September 5, 1949, and February 17, 1950; Figs. 1 (a) and 1 (b)) circular polarization was detectable towards the end of the outburst. The change to circular polarization occurred after a period of low intensity and was accompanied by an increase in intensity again. In all cases the observed position tended to return to that of the flare at the conclusion of the outburst.

(d) The Period of Non-Random Polarization

A second rise in intensity following on the first surge with polarization usually circular but sometimes elliptical or linear as distinct from the random polarization of the first surge, is so common, particularly in the outbursts of high intensity and long duration, that it may well be considered a normal second stage in an outburst. Its apparent absence in smaller outbursts may be because the intensity of this stage on such occasions is too low to be detectable. We have already pointed out that such a stage was observed in two of the four outbursts studied from their beginning and it was very marked in the other two outbursts studied (February 22 and August 5; Figs. 1 (c) and 1 (e)) for which we have, unfortunately, no interferometer measurements during the first stage. An interesting feature of these two outbursts is that during the second stage the polarization in each case became linear for a period.

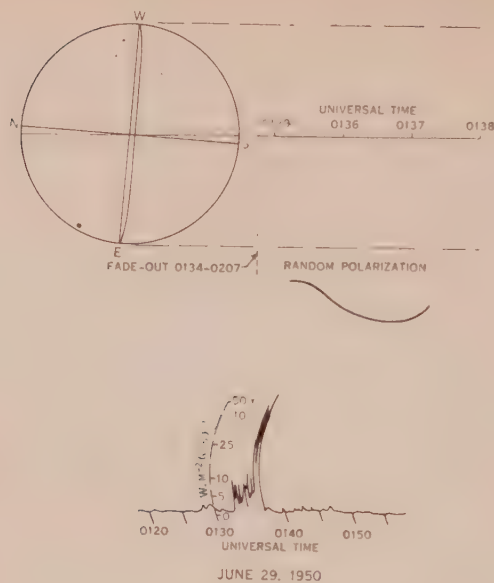


Fig. 1 (d)

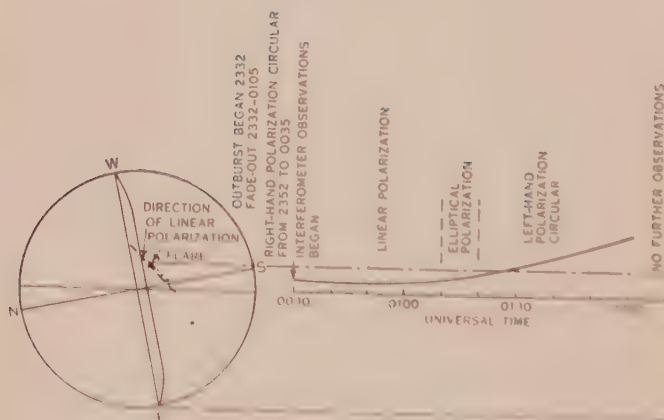


Fig. 1 (e)

The outburst of February 21-22, 1950 (Fig. 1 (c)) occurred during one of the noise storms reported in Part II of this series (Payne-Scott and Little 1951). This storm was initiated by the outburst of February 17 (Fig. 1 (b)), and at the time of the second outburst, storm radiation had been present for some time, of fairly low intensity and showing right-handed circular polarization. The initial large increase (from 2340 to about 2350 U.T.) was off-scale on the intensity record, so that the polarization is uncertain, though it was probably random. After this increase the level fell rapidly, till there remained only the previous low-intensity radiation with right-handed circular polarization; it then began

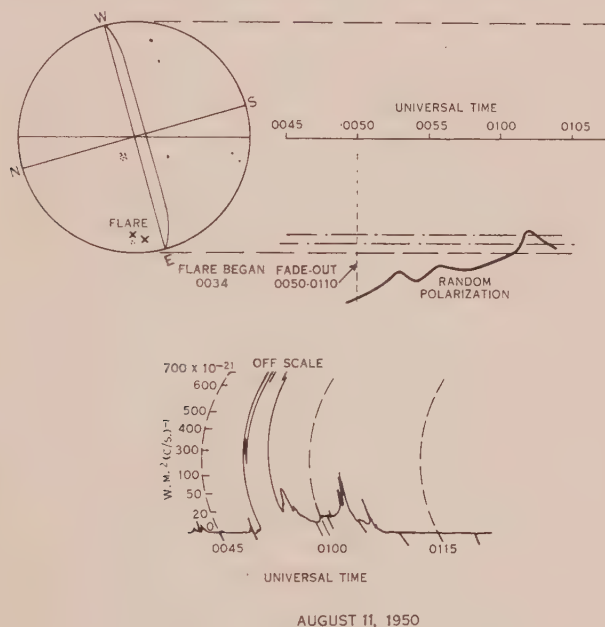


Fig. 1 (f)

to increase again, the polarization changing at the same time, so that by the time it reached the second peak the polarization was predominantly left-handed. At about this time interferometer measurements commenced, and the apparent position of the source was found to be in the area in which the flare was occurring, and moving towards the centre of the Sun. By the time the polarization was well established as left-handed the position had stabilized slightly closer to the Sun's centre than the flare, and remained there for about 20 minutes, but then began to move rapidly towards the Sun's limb. At the same time the polarization changed rapidly through elliptical to linear, then back again through elliptical to left-handed circular as the source moved back into the flare region. For the rest of the observing period the apparent position of the source remained close to this region, except for the movements between about 0200 and 0210 U.T. which will be discussed later. The direction of polarization, during the period when it was linear, was approximately along the axis of the spot-group.

The outburst of August 5, 1950 (Fig. 1 (e)) followed a similar course, except that in this case the sense of rotation of the polarization, after the period of linear polarization, was opposite to that before. This time the first large increase appears to have been an admixture of random and some right-handed circular polarization. The second increase showed marked right-handed circular polarization, changing at about 0035 U.T. to linear polarization. Interferometer observations began just after this, and showed the apparent position of the source to be close to that of the flare, and the direction of the linear polarization to be again approximately along the axis of the spot-group. At about 0110 U.T. the polarization began to change to left-handed circular, and at the same time the apparent position began to move westwards across the Sun's disk. The sense of rotation remained left-handed for the rest of the day, but no position measurements were made after 0200 U.T.

(e) *Non-Selective Fading*

The other point of interest in these interferometer measurements is the rapid oscillation in position which occurred on February 22, 1950, between about 0155 and 0215 U.T. (Fig. 1 (c)). These oscillations coincided with a remarkable fluctuation in intensity (Fig. 2) which was even more marked on the other recording frequency, 62 Mc/s. Such fluctuations have been observed on a few occasions previously by one of the authors (Payne-Scott) and have been given the name "non-selective fading" (Pawsey 1950). The characteristics of this type of intensity variation are that the variations are down as well as up, as distinct from the usual "storm bursts", which are increases above a base level; that the lower the frequency, the more marked the effect; and that the variations on different frequencies are fairly well correlated, as distinct from the normal storm bursts, which show little correlation even between closely adjacent frequencies. In this particular case the correlation coefficient between the intensities on 98 and 62 Mc/s., measured at one-minute intervals from 0200 to 0220 U.T. inclusive, was 0.43. The chance of such a correlation occurring by accident is only 1 in 20, so that the correlation is significant. The name "non-selective fading" was chosen because the variations resemble those occurring in the carrier of a distant station received via the ionosphere, but differ from them by their marked similarity on different frequencies. The slow, fairly regular, period of the variations seen here—about 2 minutes—is typical, the periods observed varying from about $\frac{1}{2}$ to 2 minutes. Also common to this occurrence, as well as to most similar occurrences observed before, is the fact that the associated flare occurred on the western half of the solar disk, and that it was followed a day or so later by a sudden-commencement magnetic storm at 1041 U.T. on February 23. There appears to be some correlation between the variations in apparent position and in intensity on February 22, but there are not enough interferometer measurements to be certain of this.

The only sudden-commencement magnetic storm that can with certainty be associated with the outbursts described here is the storm on February 23. A sudden-commencement storm was also reported on August 7 at 1055 U.T.,

and this may be connected with the flare on August 5, although the time delay is much longer than usual. No fading was observed on this occasion.

IV. INTERPRETATION

As a result of the interferometer measurements of position and polarization, we conclude that the apparent position of the source of the outburst is initially very close to that of the visual flare, but slightly further towards the solar limb; this apparent position changes rapidly, the movement usually being toward

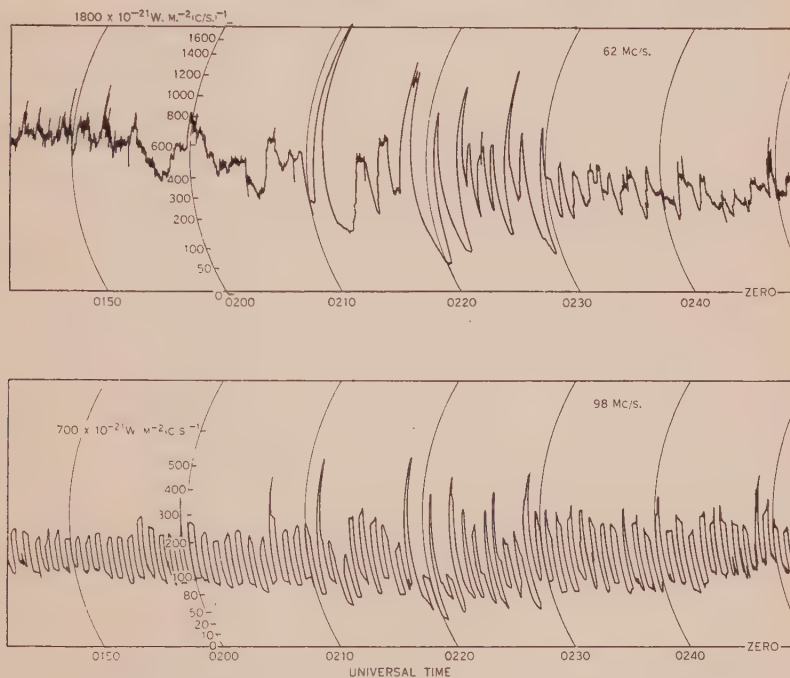


Fig. 2.—Non-selective fading on February 22, 1950; 98 Mc/s. record switched from L.H. to R.H. circular polarization in alternate half-minutes.

the solar limb; the initial polarization is random and random polarization usually persists throughout this stage of rapid movement; the apparent position slowly returns to the original place, the intensity of the radiation falling to a very low level; frequently a second increase in intensity occurs, and in this stage the polarization is elliptical, commonly circular but sometimes linear. The other observation of interest concerns the position movements during the occurrence of non-selective fading.

The close proximity between the initial position observed for the radio source during an outburst and that of the associated flare confirms the view that they are different aspects of the same disturbance. The observed apparent displacement of the outburst towards the solar limb relative to the flare is to be expected, as the lowest level in the solar atmosphere from which radiation can emerge at 97 Mc/s. is well above the photosphere, and hence (if the outburst

originates vertically above the flare) the projected positions of flare and outburst on the solar disk will show such a displacement. If we assume that the difference between the apparent positions of outburst and flare is due solely to this cause, we can calculate the height of origin of the outburst radiation. The values so calculated are given in Table 2. A range is given to allow for a possible error of ± 2 minutes of arc in the outburst measurements. The values for June 29 and August 11, 1950 (Figs. 1 (*d*) and 1 (*f*)) are, as would be expected, higher than the others, as in these cases the flare or spot-group was near the solar limb, and thus the radiation reaching the Earth travelled at a very oblique angle through the solar atmosphere; it is shown in Part II of this series that under these circumstances the apparent height of the radiating source increases. There are not sufficient measurements to give an accurate estimate of the height of origin of outburst radiation, but it seems most likely that the height for normal

TABLE 2
HEIGHTS AND VELOCITIES OF OUTBURSTS

Date	5.ix.49	17.ii.50	22.ii.50	29.vi.50	5.viii.50	11.viii.50
Height of source at beginning of outburst (photospheric radii above photosphere) ..	0-0.6	0-0.3	—	0.4-0.7	—	0.6-0.9
Greatest height of source during outburst (photospheric radii above photosphere)	4.2	1.8	—	1.0	—	0.5
Radial velocity outwards (km./sec.) from position measurements on outburst	1500	600	—	3000	—	—500
Radial velocity outwards (km./sec.) from time delay ..	1400-7000	700-3500	1800- ∞	3000- ∞	3500- ∞	400-700

incidence is about 0.3 photospheric radii, i.e. about 2×10^5 km. above the Sun's visible disk. This is about three times as high as the level for which the refractive index of the solar atmosphere is zero in the absence of a magnetic field, from calculations based on the usually assumed electron-density distribution.

The further displacement in the apparent position of the outburst toward the solar limb as it progresses can also be interpreted as a further increase in height of the source. If we assume that the source moves out radially, then from the observed positions in Figure 1 we calculate the values given in Table 2 for the velocity of the source and also for the greatest height reached.

Such an outward movement suggests that the agency initiating the outburst is itself moving physically outward through the corona. If we make the further assumption that this agency arises near the photosphere at the same time as

the visible light seen in the associated flare and the ultraviolet radiation responsible for the radio fade-out, we can also calculate a velocity for the agency from the time lag between outburst and either flare or fade-out. Table 2 shows the velocities calculated on this basis, assuming the initial height given in the first row, or an initial height of 0.3 photospheric radii when no position measurements were made at the start of the outburst. The accuracy of these measurements is very low, as the time of onset of flare or fade-out is not known to better than a minute, and hence a velocity range is shown. When possible the time of onset of the flare has been used, rather than that of the fade-out.

The velocities calculated from the time lag between the outburst and flare or fade-out are of the same order as the more accurate values calculated from the position measurements, and both point to a disturbance moving, usually outward, with a velocity of the order of 500 to 3000 km./sec. Now we know of two solar disturbances possessing such velocities. The material in eruptive prominences possesses velocities in the corona of up to 500 km./sec. The streams of ionized particles assumed to be emitted at the time of flares and to escape sometimes from the solar atmosphere and reach the Earth, giving rise to sudden-commencement magnetic storms, also have velocities of the right order. The mean velocity of these particles on the Sun-Earth path, calculated from the time delay between flare and storm, is about 1600 km./sec., agreeing well with the velocities observed during outbursts. The velocities calculated from position measurements are rather higher than those observed by Wild (1950) from measurements on the drift of spectral features with time during several outbursts in the frequency range 70–130 Mc/s. He assumes that the time differences are due to an exciting agent travelling out radially through the corona and producing radiation of each frequency at the level in the corona where the refractive index (in the absence of a magnetic field) is zero for that particular frequency. On this basis he obtains for the three outbursts which he observed an average speed for the exciting agency ranging from 320 km./sec. at 110 Mc/s. to 650 km./sec. at 70 Mc/s. There is some evidence from radio-frequency measurements that the coronal electron densities given by the conventional distribution need to be increased, at least during periods near the sunspot maximum, by a factor of perhaps 2 (see for example, Waldmeier 1951). It is pointed out in Part II of this series that such an increase is also required to explain our observations on noise storms. With such a change in distribution, the initial height of outburst radiation as observed by us would be approximately that of zero refractive index for 100 Mc/s., instead of considerably higher. The spacing between the levels of origin of different frequencies, on the assumption made by Wild, would also increase, and consequently his calculated velocities would increase, coming then into good agreement with those deduced here.

The data on the outburst which apparently moved inward from the beginning (August 11, 1950; Fig. 1 (*f*)) suggest that in this case the exciting agency was in rapid motion towards the photosphere. It could have been due to a stream of corpuscles falling back into the Sun after first being ejected without the production of an outburst. This might also explain the long delay between the onset of the flare and that of the outburst on this occasion.

The observed lack of polarization in the initial stages of an outburst, compared with the circular polarization observed during noise storms, suggests that the outburst radiation arises in and passes through a region free of magnetic fields. This is a little difficult to visualize, particularly as on some occasions, e.g. in the outburst of February 22 (Fig. 1 (c)), a noise storm was in progress before the outburst began, the polarization of the radiation being circular, so that presumably a magnetic field was already established in the region in which the outburst arose. It is possible that in the process of production of outburst radiation the proportions of ordinary and extraordinary radiation are almost exactly balanced, but this seems unlikely, as the proportion of circular polarization observed in the first stage of an outburst is usually less than 5 per cent.

The circularly-polarized radiation of much lower intensity often observed after an outburst seems to be of the same nature as storm radiation and to owe its polarization to fields present in the solar atmosphere and not to fields in some other region on the Sun-Earth path. This is confirmed by the fact that it appears to follow the polarization rule deduced in Part II of this series for storm radiation, that the direction of rotation of polarization is usually left-handed when the largest spot in the associated spot-group is a north-seeking magnetic pole and right-handed when it is a south pole, and that the observed shifts in position of radiation from above one region in a spot-group to another were accompanied by appropriate changes in polarization. Thus on February 21-22 (Fig. 1 (c)) the apparent position of the source moved from above the north pole* of the spot-group towards the centre of the group as the polarization changed from left-handed circular to linear, and then moved back as the polarization returned to left-handed; on August 5 (Fig. 1 (c)) the change in position from above the centre of the group towards the north pole of the group was again accompanied by a change from linear to left-handed polarization; since the spot-groups are in opposite hemispheres the movements are in opposite directions, in the first case to the east and in the second to the west. In the only other case in which a change of polarization was observed, on February 17, 1950 (Fig. 1 (b)), the change of polarization from right-handed to left-handed was accompanied by an apparent position change from above a south to above a north pole, as given by the above rule; however, in this case it is likely that the major part of the apparent position change is due to a change in height of the source rather than to a tangential movement.

It has been pointed out in Part II of this series that the observed polarization rule, in so far as it concerns right-handed and left-handed circular polarization, can most easily be explained on the assumption that it is the "ordinary"†

* In the present solar cycle the leading spots in a group are south-seeking poles and the followers are north-seeking poles in the northern hemisphere, while the polarities are reversed in the southern hemisphere.

† The "ordinary" ray is the term commonly used for the ray for which the electron density for zero refractive index is unaffected by the presence of a magnetic field when propagation is quasi-transverse.

ray of magneto-ionic theory that escapes. The direction of the observed linear polarization, with the E -vector approximately along the axis of the spot-group, is also that which would be expected for an ordinary ray propagated in the region in which the lines of magnetic force are parallel to the line joining the two spots, while the extraordinary ray would be polarized at right angles to this line.

The frequent occurrence of circularly-polarized radiation after an outburst when none has been present before, or the increase in intensity of an existing storm that often follows an outburst, suggests that the passage of the agency initiating the outburst in some way modifies conditions in the corona so as to make the production of storm radiation possible. Since the occurrence of storm radiation is so intimately connected with the presence of a magnetic field, it is possible that the growth of the spot-field in the corona is stimulated during an outburst.

Refraction as a possible cause of the observed position changes has not so far been considered, for, as pointed out in Part II of this series, special refracting regions of particular shape and density would have to be assumed to produce bending towards the solar limb, while the observed movements can be more simply explained as due to an actual movement of the emitting source.

V. CONCLUSIONS

The evidence presented here establishes that outbursts of radio noise are another manifestation of the solar disturbances already known to produce optical flares, radio fade-outs, and terrestrial magnetic storms. The apparent changes in position of the source of outburst radiation can most readily be interpreted as due to the passage through the solar corona of an exciting agent moving with a speed of from 500 to 3000 km./sec. It seems likely that normally the exciting agent is the corpuscular stream which is assumed to cause terrestrial magnetic storms. The motion of the exciting agent is usually outward in the corona, though on one occasion the evidence points to an inward motion. The speeds measured are of the same order as, but rather greater than, those determined by Wild (1950) from spectrum analysis measurements; it is possible that Wild's values may have been depressed through the use of levels of origin which are lower than those suggested by evidence both in this paper and from other sources. The interferometer measurements suggest that the level of origin of the outburst radiation is initially at or above the level of zero refractive index and eventually reaches a height of about two photospheric radii above the photosphere. The absence of coherent polarization during the early stages of an outburst is difficult to explain satisfactorily.

The radiation with coherent polarization often appearing later seems to be identical with the storm radiation discussed in Part II of this series (Payne-Scott and Little 1951) and follows the same rule for the direction of polarization. Thus the radiation arising above a south magnetic pole shows right-handed circular polarization, that above a north pole is left-handed, and that above the region between the poles of a bipolar group is linear, with the E -vector directed along the axis of the group. It is suggested that the appearance of coherent

polarization after an outburst may be due to the development of coronal magnetic fields during the outburst.

The peculiar type of variation known as "non-selective fading" was encountered on one occasion during the measurements recorded here. However, the authors do not wish to make any definite suggestions about the origin of this fading without first considering the results of more detailed observations.

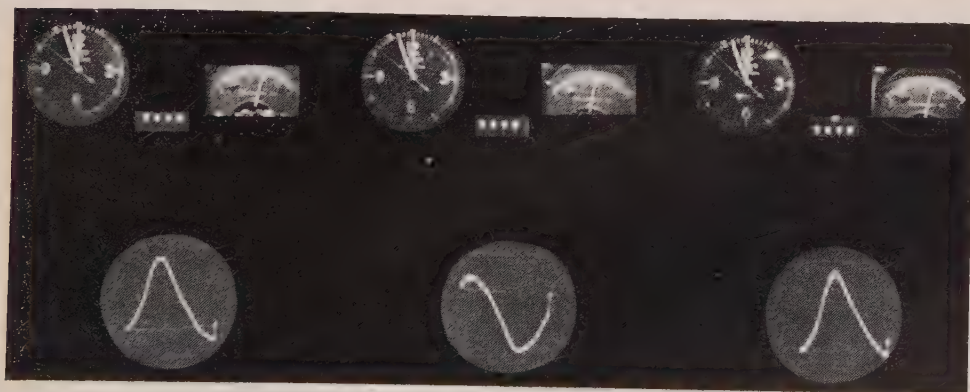
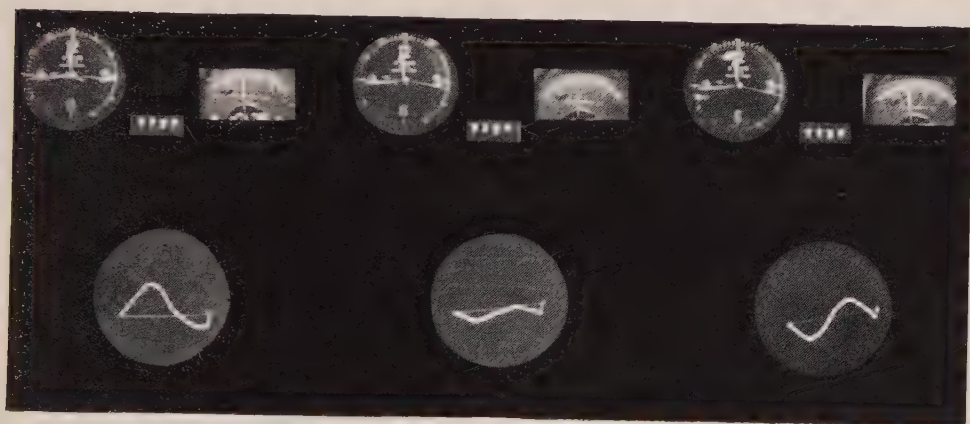
VI. ACKNOWLEDGMENT

The authors would like to acknowledge the encouragement and critical advice of Dr. J. L. Pawsey throughout this work.

VII. REFERENCES

- ALLEN, C. W. (1947).—*Mon. Not. R. Astr. Soc.* **107**: 386-96.
HEY, J. S. (1946).—*Nature* **157**: 47-8.
LITTLE, A. G., and PAYNE-SCOTT, RUBY (1951).—*Aust. J. Sci. Res. A* **4**: 489.
PAWSEY, J. L. (1950).—*Proc. Instn. Elect. Engrs.* **97**: 290-310.
PAYNE-SCOTT, RUBY (1949).—*Aust. J. Sci. Res. A* **2**: 214-27.
PAYNE-SCOTT, RUBY, and LITTLE, A. G. (1951).—*Aust. J. Sci. Res. A* **4**: 508.
WALDMEIER, M. (1951).—*Naturwissenschaften* **38**: 1-4.
WILD, J. P. (1950).—*Aust. J. Sci. Res. A* **3**: 399-408.

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. III



Interferometer records for February 17, 1950, showing change from random to circular polarization during course of outburst.

THE HIGH LATITUDE EAST-WEST ASYMMETRY OF COSMIC RAYS*

By D. W. P. BURBURY† and K. B. FENTON‡

[*Manuscript received October 2, 1951*]

Summary

Measurements of the east-west asymmetry of cosmic rays at sea-level have been made at Hobart and Macquarie Island in geomagnetic latitudes 51.7° and 60.7° S. respectively. The values obtained have been found to agree satisfactorily with values calculated using a revised form of Johnson's theory.

I. INTRODUCTION

The theory of the geomagnetic effects of cosmic rays has been discussed by several workers (Rossi 1930 ; Lemaitre and Vallarta 1932, 1933 ; Swann 1933 ; Johnson 1938), and a satisfactory qualitative explanation has been given for the latitude effect and the east-west asymmetry in low latitudes. This theory is based on the effect of the Earth's dipole field on charged primary cosmic ray particles. In order to explain the absence of any variation of the cosmic ray intensity near sea-level between the so-called "knee" of the latitude effect at about 40° geomagnetic latitude and the poles, it is necessary to assume that field sensitive primary particles do not contribute to the sea-level radiation in these high latitudes. This implies isotropic primaries for the generation of the sea-level radiation and hence zero sea-level asymmetry in regions above the knee.

A small east-west asymmetry has, however, been shown to exist in geomagnetic latitudes of 50° and over by Johnson (1941), Seidl (1941), Fenton and Burbury (1948), and others. This was explained by Johnson (1941) by considering the deflection in the Earth's field of secondary mesons during their passage through the atmosphere. He showed that the deflection, although small, was sufficient to produce an asymmetry of the right order. The experimental results available to him were not sufficient to test the theory adequately, because of their large probable errors. Measurements of the asymmetry have been made at Hobart (51.7° S. geomagnetic latitude) and Macquarie Island (60.7° S. geomagnetic latitude) to supplement those used by Johnson. At the same time the calculation of the asymmetry has been repeated using numerical

* Most of this paper formed part of the first author's Ph.D. thesis which has been accepted by the University of Tasmania. The other author (K.B.F.) was engaged in the experimental work on Macquarie Island. On his return to Tasmania (May 1951) it was found that he had undertaken independently a revision of Johnson's theory. Since it differed only in a few details from that given here, this paper has been written under both names.

† C.S.I.R.O., attached to Physics Department, University of Tasmania.

‡ Australian National Antarctic Research Expedition, attached to Physics Department, University of Tasmania.

methods to make possible the use of experimental data which have become available since Johnson's calculations were made.

II. THEORY OF THE ASYMMETRY

(a) *Introduction*

It is assumed that the part of the primary radiation which contributes to the intensity at sea-level is isotropic above the atmosphere in latitudes above the knee of the latitude effect. This theory treats the asymmetry in the radiation which arises because the secondary mesons, which form nearly all of the penetrating component at sea-level, undergo slight deflections in the Earth's magnetic field.

If the magnetic field were absent, absorption effects in the atmosphere would result in a directional distribution of the intensity at sea-level which would be symmetrical about the vertical, the intensity being a function of the zenith angle only. Since, as will be shown later, the change in direction of the path of a meson of even very low energy is not more than a few degrees, the effect of the deflections will be to slightly upset the symmetrical distribution which would be found in the absence of the magnetic field. Because of the direction of the deflections the intensity of positive particles which would be found at a certain zenith angle in the east-west vertical plane in the symmetrical distribution will be found a few degrees to the west, while for negative particles the displacement will be to the east. Since there are more positive than negative mesons there is an asymmetry in the total radiation.

The general lines of Johnson's theory have been followed here, i.e. the deflections suffered by mesons with different final energies and directions have been calculated first, and then the asymmetry which arises because of these deflections. The main difference between the treatment given here and that given by Johnson lies in the use of graphs and numerical integration as an aid to calculation. This has enabled empirically determined data to be used without the difficulty of fitting analytical expressions, which would lead to a more complicated treatment. The errors in the values of the asymmetry calculated by these methods may be expected to be considerably smaller than the probable errors of the experimental results with which they are to be compared.

Only mesons travelling in the east-west vertical plane are considered. The curvature $1/\rho$ of a particle's trajectory at any point is a function of the energy of the particle when its mass and charge and the value of the horizontal component of the magnetic field are known. Since the total deflection suffered by the mesons during their passage through the atmosphere is no more than a few degrees, it has been considered a sufficiently good approximation to take the integral of the quantity $1/\rho$ along a straight path as the deflection of a particle whose final direction lies along that path. By expressing the final energy in terms of the residual range in air and using energy-range relations and pressure-altitude tables given by Montgomery (1949, pp. 347, 353), it is possible to obtain the value of $1/\rho$ at equally spaced points along straight lines inclined at 15, 30,

45, and 60° to the vertical. These are then integrated by Simpson's method to obtain the value of

$$\delta = \int_0^{s_1} \frac{ds}{\rho},$$

where s_1 has been chosen to correspond with the average height of formation of the mesons. The values of δ in radians for Hobart are given in Table 1 as a

TABLE 1

DEFLECTION δ AS A FUNCTION OF ZENITH ANGLE AND FINAL RESIDUAL RANGE IN AIR (IN RADIAN)

Residual range (g. cm. ²)	50	100	200	400	800	1500	3000	6000
Final energy m_0c^2	1.3	2.2	4.0	8.0	16.0	32.5	70	140*
Zenith angle								
15°	0.085	0.075	0.062	0.046	0.030	0.019	0.011	0.005*
30°	0.090	0.078	0.065	0.048	0.034	0.021	0.012	0.006*
45°	0.092	0.082	0.069	0.054	0.037	0.023	0.014	0.007*
60°	0.104	0.089	0.076	0.060	0.044	0.030	0.019	0.010*

* Approximately.

function of zenith angle of arrival and final residual range in air. The values of the horizontal component of the Earth's magnetic field have been taken as 0.19 oersteds for Hobart and 0.13 oersteds for Macquarie Island.

TABLE 2

COMPARISON OF DEFLECTION CALCULATED BY JOHNSON'S AND
NUMERICAL METHODS (IN RADIAN)

For direction of arrival at 45° to vertical

Final Energy (unit 100 MeV.)	Johnson's Method	Numerical Method
2.5	0.0278	0.0362
5	0.0202	0.0316
10	0.0128	0.0222
20	0.0068	0.0153
40	0.0030	0.0101
80	0.0011	0.0068

The deflections calculated by the numerical method are slightly greater than those calculated by Johnson's method, as is shown in Table 2, where the deflections of particles of different final energies arriving at an angle of 45° to the vertical calculated by the two methods for the latitude of Macquarie Island are

given. The difference is more marked for higher energy particles. This is attributed to the fact that we have used the expression

$$\delta = \int_0^{s_1} \frac{ds}{\rho},$$

where Johnson used

$$\delta = \lim_{s_1 \rightarrow \infty} \int_0^{s_1} \left(\frac{1}{\rho} - \frac{1}{\rho_0} \right) ds,$$

where ρ_0 is the radius of curvature of the path of the particle before it is slowed down by the atmosphere. Since the former expression is more in keeping with the idea that the majority of the μ -mesons originate in a fairly well-defined layer near the 100 mb. level of the atmosphere, it is considered that it gives better values of δ for use in estimating the asymmetry.

(b) Calculation of the Asymmetry

The asymmetry which arises due to the deflections has been calculated by the method used by Johnson (1941). It is assumed that in the absence of the magnetic field the directional distribution of the radiation would be symmetrical about the vertical, the intensity at a zenith angle θ being given by

$$j(\theta) = j(0) \cos^\gamma \theta,$$

$j(0)$ being the vertical intensity. The value of γ has been discussed by several authors (Kraushaar 1949; Rogozinski and Voisin 1949; Zar and Shamos 1950). Experimental results indicate that the intensity of slow mesons with energy about 2×10^8 eV. is approximately proportional to $\cos^3 \theta$ but for energies of 5.5×10^8 eV. and greater, the variation follows $\cos^{2.2} \theta$ fairly closely. Since the assumption of a $\cos^{2.2} \theta$ variation introduces an error for only a small portion of the meson spectrum, the value $\gamma = 2.2$ has been used here.

It has been assumed (following Johnson) that the effect of the small deflections is to displace the intensity of any particular energy band of mesons by an amount equal to their deflection. For the positive or negative mesons taken alone this leads to an asymmetry

$$A = 4.4 \bar{\delta} \tan \theta,$$

where $\bar{\delta}$ is the average value of the deflection, the average being taken over the energy spectrum with a lower limit equal to the lowest energy with which a meson can penetrate the amount of absorber used in the apparatus with which the asymmetry is measured (cf. Johnson 1941). The differential range spectrum given by Rossi (1948) has been used in the calculation of $\bar{\delta}$, and Table 3 gives the values obtained. Table 4 gives the corresponding values of the asymmetry of the purely positive or negative meson component, assuming these to be numerically equal but of opposite sign, the intensity of the positive particles being greater in the west.

If we denote by W^+ , E^+ , W^- , and E^- , the intensities of the positive and negative particles at an angle θ to the west and east of the vertical, and use

$$P = \frac{2(W^+ + E^+ - W^- - E^-)}{W^+ + E^+ + W^- + E^-}$$

as a measure of the positive excess, it can easily be shown that the asymmetry of the total radiation is given by

$$\alpha = \frac{1}{2}Ap = 2 \cdot 2p\bar{\delta} \tan \theta.$$

TABLE 3
AVERAGE DEFLECTION $\bar{\delta}$ AS A FUNCTION OF ZENITH ANGLE AND LOW
ENERGY CUT-OFF FOR HOBART (IN RADIANs)

Zenith angle ..	15°	30°	45°	60°
Low energy cut-off ($2 \cdot 2 \times 10^8$ eV.) ..	0.031	0.034	0.037	0.043
Low energy cut-off ($4 \cdot 0 \times 10^8$ eV.) ..	0.028	0.031	0.034	0.040

The ratio of the number of positive to negative particles determined by Owen and Wilson (1949) has been used here. The value given is 1.268 ± 0.023 , and gives a value of 0.236 for p . This value is not inconsistent with later

TABLE 4
ASYMMETRY OF PURELY POSITIVE OR NEGATIVE COMPONENT AS A
FUNCTION OF ZENITH ANGLE AND LOW ENERGY CUT-OFF FOR HOBART

Zenith angle ..	15°	30°	45°	60°
Low energy cut-off ($2 \cdot 2 \times 10^8$ eV.) ..	0.0365	0.0864	0.1628	0.3277
Low energy cut-off ($4 \cdot 0 \times 10^8$ eV.) ..	0.0330	0.0788	0.1496	0.3048

determinations (Owen and Wilson 1951; Caro, Parry, and Rathgeber 1951) at least within the limits of accuracy required for comparison with the experimental results. It must be stressed that the data used in the determination of

TABLE 5
ASYMMETRY OF TOTAL RADIATION AS A FUNCTION OF ZENITH ANGLE
AND LOW ENERGY CUT-OFF FOR HOBART

Zenith angle ..	15°	30°	45°	60°
Low energy cut-off ($2 \cdot 2 \times 10^8$ eV.) ..	0.0043	0.0102	0.0192	0.0387
Low energy cut-off ($4 \cdot 0 \times 10^8$ eV.) ..	0.0039	0.0092	0.0177	0.0360

the asymmetry are obtained from experiments conducted on the radiation at or near the vertical in the absence of information for the inclined directions, and it is to be expected that the calculated values should be more reliable for the smaller zenith angles.

The calculated values of the asymmetry of the total penetrating radiation are given in Table 5 for different zenith angles and lower energy limits corresponding to the two thicknesses of lead used in the apparatus.

III. HOBART APPARATUS

An automatic apparatus has been used in measuring the asymmetry. This consists of two Geiger counter telescopes arranged to point at the same zenith angle either side of the vertical in the east-west plane. The telescopes are mounted on a turntable which can be rotated through 180° every hour to interchange the positions of the two telescopes. Runs of 1 hour are made in each position, and the number of counts in each hour obtained by photographing at the beginning and end of each run mechanical recorders connected to the telescopes.

The turntable base is provided with three levelling screws and the axis of rotation of the turntable was set vertical by the following procedure at the commencement of the experiments. A plumb-line of fine wire was hung from a stand on the turntable and a telemicroscope was mounted on the turntable and focused on the wire near its lower end. The turntable was levelled until the movement of the wire in the field of the telemicroscope was less than its own diameter during a rotation of the turntable. A simple calculation then showed that the axis of rotation was set within 0.0038° or about $\frac{1}{4}$ minute of the vertical. The turntable assembly which is sturdily constructed of cast iron is mounted on a concrete block 2 ft. square sunk several feet into the ground to ensure that this setting is maintained.

The possible error in the azimuth setting of the apparatus was estimated to be of the order of 1° , but the determination of the asymmetry does not depend on this setting so much as on the reproducibility of the zenith setting after each rotation.

The automatic operation of the apparatus is controlled by a timing mechanism actuated by a segmented armature type of synchronous motor. The lengths of the hourly runs are thus controlled with the accuracy of an electric clock. The frequency of the mains supply is regulated by the Hydro-electric Commission so that an electric clock seldom varies from the correct time by more than 10 seconds and the error over each day is adjusted to zero. The use of the double telescope apparatus minimizes the effect of such fluctuations as they affect one telescope opposite the other.

Each telescope consists of two banks of Geiger counters mounted in a strong framework so that lead blocks may be placed between them. Runs have been made at several zenith angles with 0, 12, and 24 cm. of lead between the counter trays.

The telescopes were enlarged about half way through the experiments to increase the counting rate so that less time was needed to obtain a significant value of the asymmetry at any one setting. The first telescopes each consisted of two sets of three copper in glass counters with 6 by 1 in. cathodes separated by 12.5 in. and, taking the sensitive area of the counters to coincide with the physical size of the cathodes, this gave a sensitive solid angle extending up to 13.5° either side of the zenith angle setting and 25.5° either side of the azimuth setting. When the telescopes were enlarged, counters of the external cathode type were installed. The dimensions of the trays were made 20 by 20 cm. (top tray) and 20 by 15 cm. (bottom tray), and their separation 75 cm., so that

the sensitive solid angle extended about 13° either side of the zenith and 15° either side of the azimuth setting. It was found that the external cathode counters had a high background counting rate, due probably to the presence of K^{40} in the soda glass used, and this increased the accidental rate of the telescopes to a value where it had to be corrected in the computation of the results.

IV. RESULTS OF HOBART EXPERIMENTS

At the end of a period of running at a certain zenith setting and lead thickness, the rates A_W , A_E , B_W , and B_E were calculated, the telescopes being designated A and B and the suffixes denoting the direction. The asymmetries and their probable errors were then calculated in the usual way for each telescope. Each asymmetry was weighted in proportion to the reciprocal of the square of its probable error and the weighted mean taken as the best estimate of the asymmetry (cf. Bond 1935, pp. 81, 83). This method of calculating the asymmetry was applicable where the apparatus was used at the same setting during more than one period.

It was necessary to apply a correction to the results obtained with the enlarged telescopes due to the increased accidental rate experienced when the external cathode counters were used. This accidental rate was determined separately to be about one count per minute. In view of the difficulty of determining the accidental rate for a twofold coincidence telescope this rate was taken to be the same for all settings.

Table 6 gives the data obtained in the experiments, and Table 7 the weighted means of the asymmetry at the different zenith settings and lead thicknesses used.

An analysis of variance has been carried out on 63 days' results obtained during a run at 45° with 12 cm. of lead in the telescopes. Hourly values of the asymmetry were classified into daily groups, and the estimates of variance between and within days compared. The result showed no significant difference between the two estimates. This indicates that the value of the asymmetry does not depend strongly on any external factors which vary from day to day. A dependence of the asymmetry on, for example, the barometric pressure would increase the day-to-day fluctuation and would lead to a significant result in the above test.

V. RESULTS OF THE MACQUARIE ISLAND EXPERIMENTS

Preliminary results of the Macquarie Island experiments during 1950-51 are given in Table 7. It is felt that, although the results have not yet been thoroughly examined, the figures quoted here are not greatly in error and will serve to indicate the way in which the asymmetry varies with latitude.

The apparatus used at Macquarie Island did not differ in principle from that at Hobart. Details of the equipment will be published elsewhere at a later date.

VI. COMPARISON OF THEORY AND EXPERIMENT

In Figure 1 the asymmetry is plotted against the thickness of lead absorber used in the apparatus for zenith-angle settings of 45 and 60° . The fact that the

asymmetry is not greatly affected by the absorber supports the contention that the asymmetry is due to the deflections of the mesons, which make up the

TABLE 6

DATA OBTAINED FROM HOBART EXPERIMENTS

The total number of counts recorded and the total recording time in hours (upper and lower numbers respectively) are given for each telescope

Description of Run	Telescope A		Telescope B		Set Used*
	West	East	West	East	
60°, no lead, 8.xii.48-28.ii.49 ..	58610 574	66164 661	64885 668	55336 582	D.T.1
60°, no lead, 30.vi.48-5.viii.48 ..	12667 119	12724 121	23420 239	22885 233	D.T.1
60°, no lead, 24.vii.50-17.viii.50 ..	63864 239	63732 239	69682 257	68274 258	D.T.M.
60°, 12 cm. lead, 5.viii.48-20.xi.48 ..	33526 522	29797 478	31995 479	34054 523	D.T.1
60°, 12 cm. lead, 17.viii.50-14.ix.50 ..	40798 227	39530 226	20720 116	20346 116	D.T.M.
60°, 24 cm. lead, 28.ii.49-23.vi.49 ..	49773 790	47387 775	43291 702	44237 723	D.T.1
45°, no lead, 22.vi.50-3.vii.50 ..	54144 110	52212 108	45268 92	45676 94	D.T.M.
45°, 12 cm. lead, 14.ix.50-4.x.50 ..	65808 194	64434 194	65287 194	64810 194	D.T.M.
45°, 12 cm. lead, 8.i.51-16.ii.51 ..	136920 403	131498 395	133986 395	133267 403	D.T.M.
45°, 24 cm. lead, 3.vii.50-24.vii.50 ..	64498 199	62906 197	63770 197	63190 199	D.T.M.
30°, 12 cm. lead, 1.x.50-14.x.50 ..	33679 67	34590 69	34476 69	32914 67	D.T.M.
15°, 12 cm. lead, 20.x.50-20.xi.50 ..	118819 185	115649 181	139531 218	141164 222	D.T.M.

* D.T.1, first double telescope; D.T.M., modified double telescope.

penetrating component. The slight increase observed when 12 cm. are introduced is probably due to the elimination of a nearly symmetrical soft component which would have the effect of diluting the asymmetry in measurements with

no lead. The decrease between 12 and 24 cm. can be attributed to the elimination of low energy mesons, which because of their greater deflection would contribute more to the asymmetry than higher energy mesons.

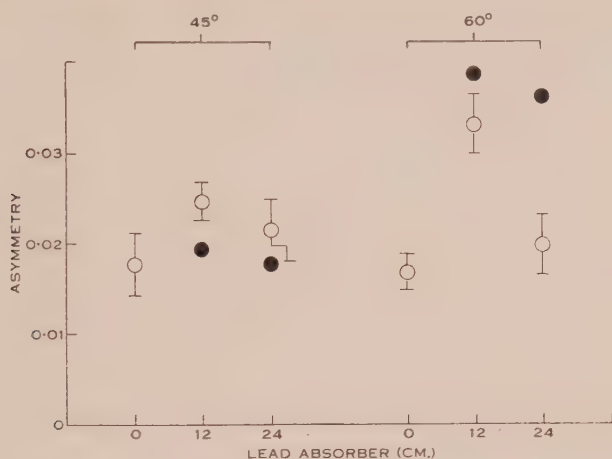


Fig. 1.—The effect of lead absorber on the asymmetry at 45 and 60° zenith angles. The results are compared with the theoretical values.

○ Hobart experimental results. ● Theoretical values.

In comparing the zenith-angle variation of the asymmetry (Fig. 2) with that given by the theory, it must be borne in mind that the calculated values are based on data which are more reliable for directions near the vertical than

TABLE 7

COMPARISON OF HOBART AND MACQUARIE ISLAND ASYMMETRY

Ratio of horizontal component of magnetic field at Hobart and Macquarie Island = $0.19/0.13 = 1.46$

Description	Hobart	Macquarie Island	H/M.I. = x^*
15°, 12 cm. lead	0.0066 ± 0.0021	0.0060 ± 0.0029	1.10
30°, 12 cm. lead	0.0113 ± 0.0042	0.0049 ± 0.0042	2.31
45°, no lead ..	0.0175 ± 0.0035	0.0104 ± 0.0037	1.68
45°, 12 cm. lead	0.0245 ± 0.0019	0.0139 ± 0.0025	1.76
45°, 24 cm. lead	0.0213 ± 0.0034	0.0107 ± 0.0053	1.99
60°, no lead ..	0.0168 ± 0.0020	0.0144 ± 0.0043	1.17
60°, 12 cm. lead	0.0303 ± 0.0032	0.0265 ± 0.0052	1.14
60°, 24 cm. lead	0.0197 ± 0.0031		
70°, 12 cm. lead		0.0277 ± 0.0110	

* $\bar{x} = 1.59 \pm 0.12$.

for inclined directions. Therefore, it may be expected that if the theory gives a good account of the phenomenon then the agreement will probably be best at the smaller zenith angles. In view of this the agreement found can be regarded

as quite satisfactory. Since the geomagnetic latitude of Macquarie Island is 60.7°S. , it should be well beyond the knee of the latitude effect. The fact that the results obtained at the Island station show a definite asymmetry gives strong support to the assumption that the asymmetry in these latitudes is a secondary radiation effect. Furthermore, the values obtained at the two stations are nearly proportional to the strength of the horizontal magnetic field, which would be expected if the theory given in the text were valid. The ratio

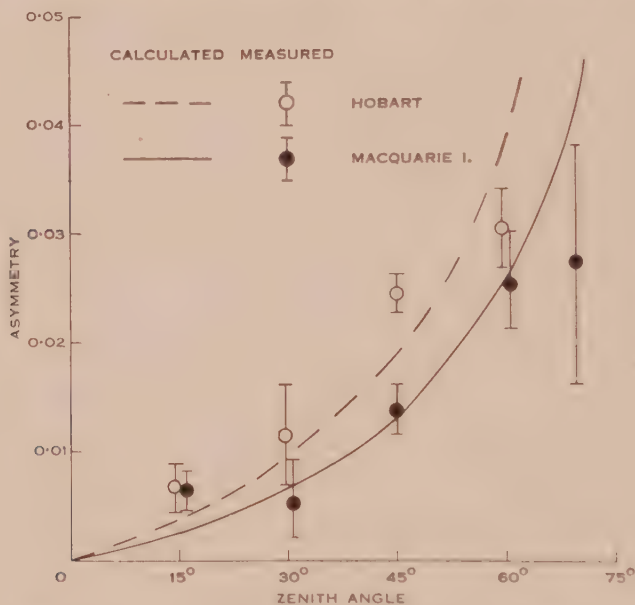


Fig. 2.—Experimental values obtained at Hobart and at Macquarie Island using 12 cm. of lead absorber in the telescopes. The curves show the corresponding theoretical values. Those for Macquarie Island were obtained by reducing those for Hobart in the ratio of the horizontal component of the Earth's magnetic field.

of the Hobart to Macquarie Island values averaged over all the settings used at both stations is found to be 1.59 ± 0.12 , compared with the ratio of the values of the horizontal component of the magnetic field, 1.46.

To obtain theoretical values with which to compare the Macquarie Island results, the values for Hobart have been reduced in proportion to the horizontal component of the magnetic field, for since the curvature of the path of a meson depends linearly on the magnetic field, so does the calculated asymmetry if all the other quantities used in the theory are assumed to be the same at both stations. Figure 2 also shows the values of the asymmetry measured at Macquarie Island with 12 cm. of lead in the apparatus, and these are compared with the calculated values. The points lie sufficiently close to the curve to show that the theory gives a good account of the asymmetry, at least up to zenith angles of between 60 and 70° .

It can be seen in both cases that the experimental results indicate a falling off of the asymmetry at the greatest zenith angles used. Without experimental determinations closer to the horizon, however, this indication cannot be given very much weight. At Hobart it was not possible to make measurements at greater zenith angles than 60° because of the prominence of Mt. Wellington in the western direction. Similar difficulties prevented the extension of the Macquarie Island experiments to any greater angle than 70° .

VII. CONCLUSION

The results obtained in the Hobart experiments have been found to agree well with values of the asymmetry predicted by a form of Johnson's theory revised to include more recent information on the energy spectrum and the positive excess of the meson component.

Comparison of the Hobart and Macquarie Island results further supports the theory by showing that the asymmetry at the two stations is proportional to the horizontal component of the Earth's magnetic field, which is predicted by the theory.

We can now state the conclusion that the asymmetry in high latitudes can be satisfactorily accounted for as a secondary radiation effect. This removes the difficulty which arises because the theory of the geomagnetic effects based on the deflection of the primary cosmic ray particles in the Earth's dipole field requires the east-west effect to be zero in latitudes where the variation of the total cosmic ray intensity with latitude ceases, as shown by Compton and Turner (1937), and by Law, McKenzie, and Rathgeber (1949), i.e. at latitudes above the knee of the latitude effect.

VIII. ACKNOWLEDGMENTS

The Hobart experiments were made possible by the assistance of C.S.I.R.O. and those at Macquarie Island by the assistance of the Australian National Antarctic Research Expedition. The authors wish to make grateful acknowledgment to these organizations. It is also a pleasure to thank Professor A. L. McAulay and Dr. A. G. Fenton of the Physics Department, University of Tasmania, for their continued interest and encouragement, and Mr. J. K. Parry, University of Melbourne, for private communication of results prior to their publication. The authors would also like to acknowledge the advice of Professor E. J. G. Pitman on the analysis of variance mentioned in Section IV, and the help of Mr. A. C. McLaren in carrying this out.

IX. REFERENCES

- BOND, W. M. (1935).—"Probability and Random Errors." (Edward Arnold and Co.: London.)
CARO, D. E., PARRY, J. K., and RATHGEBER, H. D. (1951).—*Aust. J. Sci. Res. A* **4**: 16.
COMPTON, A. H., and TURNER, R. M. (1937).—*Phys. Rev.* **52**: 799.
FENTON, A. G., and BURBURY, D. W. P. (1948).—*Phys. Rev.* **74**: 589.
JOHNSON, T. H. (1938).—*Rev. Mod. Phys.* **10**: 227.
JOHNSON, T. H. (1941).—*Phys. Rev.* **59**: 11.
KRAUSHAAR, W. L. (1949).—*Phys. Rev.* **76**: 1045.
LAW, P. G., MCKENZIE, C. D., and RATHGEBER, H. D. (1949).—*Aust. J. Sci. Res. A* **2**: 493.

- LEMAITRE, G., and VALLARTA, M. S. (1932).—*Phys. Rev.* **42**: 914.
- LEMAITRE, G., and VALLARTA, M. S. (1933).—*Phys. Rev.* **43**: 87.
- MONTGOMERY, D. J. X. (1949).—"Cosmic Ray Physics." (Princeton Univ. Press: New Jersey.)
- OWEN, B. G., and WILSON, J. G. (1949).—*Proc. Phys. Soc. A* **62**: 601.
- ROGOZINSKI, A., and VOISON, A. G. (1949). "Cosmic Radiation." Colston Res. Pap. p. 69.
(Butterworth's Scientific Publications: London.)
- ROSSI, B. (1930).—*Phys. Rev.* **36**: 606.
- ROSSI, B. (1948).—*Rev. Mod. Phys.* **20**: 545.
- SEIDL, F. G. P. (1941).—*Phys. Rev.* **59**: 7.
- SWANN, W. F. G. (1933).—*Phys. Rev.* **44**: 224.
- ZAR, JACOB L., and SHAMOS, M. H. (1950).—*Phys. Rev.* **79**: 207.

THE GROWTH OF CLOUD DROPS BY CONDENSATION

I. GENERAL CHARACTERISTICS

By P. SQUIRES*

[*Manuscript received August 7, 1951*]

Summary

Condensation in cloud is studied to demonstrate certain general characteristics, without special assumptions about the condensation nuclei or the vertical velocity. The equation of growth of a droplet is constructed, taking account of ventilation due to its falling velocity. Lifting is assumed to occur adiabatically. An approximate value is found for the supersaturation of cloud air, which under certain conditions can be computed from observable quantities—the cloud droplet spectrum and the vertical velocity. Growth of drops at small sizes is very rapid: at constant supersaturation, in a period of a few seconds, a drop in general either grows beyond $r=1\mu$, or else comes so close to equilibrium at some smaller size that growth practically ceases. When the supersaturation is moderate or large, a threshold effect appears: the size of the condensation nucleus determines whether or not the drop shall grow, but, if it does, has little influence on its final size. The definition of the size of a small condensation nucleus is reviewed and a more precise interpretation is suggested.

I. INTRODUCTION

The most satisfactory treatments of the growth of cloud droplets by condensation are those of Kraus and Smith (1949) and Howell (1949). In these papers the evolution of the spectrum of drop sizes is computed numerically for certain special distributions of nuclei and constant upcurrents. Howell took account of the heat balance of the drop. The effect of the drop's motion was neglected.

These investigators found in all cases that the resulting cloud contained a rather narrow range of drop sizes, even though the nuclei (measured by molar masses) showed a wide dispersion. This rather surprising result was explained by the behaviour of the supersaturation, which increased from zero near the cloud base to a maximum not far above, and thereafter declined steadily. If a nucleus became activated during the phase of high supersaturation, it tended to grow rapidly to a size at which the nucleus exerted little effect on the drop's growth.

The purpose of this paper is to investigate whether the neglect of the effect of ventilation of the drop due to its falling motion introduces serious errors, and to formulate the equations of growth without restriction as to velocity or nucleus spectrum, in order to reach conclusions of wider generality, and in particular to determine the general conditions suitable for the formation of large drops, which will be discussed in more detail in Part II of this series.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

Süring (1927) and Findeisen (1939) recommended that $r=100\mu$ should be accepted as a conventional upper limit of cloud drop sizes. The value is accepted here and the analysis is limited to drops up to this size.

The numerical values of quantities which depend on temperature and pressure are computed for 10 °C. and 800 mb. except where otherwise stated. A list of symbols is given in Section VI.

II. GROWTH OF A DROPLET BY CONDENSATION

(a) *Transfer of Heat and Water Vapour*

The growth of a droplet by condensation involves the transfer of both water vapour and heat, for the quantities of latent heat released during any appreciable growth by condensation are large compared with the thermal capacity of the drop, as will be shown in detail below.

The methods of dynamical similitude (see, for example, Goldstein 1938) lead to the result that heat is transferred to a ventilated spherical drop (the diameter of which is large compared with the mean free path of the air molecules) at the rate

$$4\pi kr(\theta - \theta_d)F_1(R_e, \sigma) \text{ cal. sec.}^{-1},$$

where k is the conductivity of air, r the drop radius, θ the air temperature, θ_d the temperature of the surface of the drop, $F_1(R_e, \sigma)$ is a function of R_e , the Reynolds number, and σ the Prandtl number;

$$R_e = \frac{2rV}{\nu},$$

where V is the velocity of the drop relative to the air, and ν the kinematic viscosity of air;

$$\sigma = \frac{\nu}{\nu_c},$$

where ν_c is the thermometric conductivity, or thermal diffusivity of air.

Analogously, the rate at which water vapour is transferred to a ventilated spherical drop (with a similar condition as to size) is

$$4\pi \frac{\varepsilon D}{RT} r(e_a - e_d)F_2(R_e, \sigma') \text{ g.sec.}^{-1}, \quad \dots\dots\dots (2.1)$$

where ε is the specific gravity of water vapour with respect to air, D the coefficient of diffusion of water vapour in air, R the gas constant per gram of dry air, T the absolute temperature at which the process occurs, e_a the vapour pressure of the air, and e_d the equilibrium vapour pressure of the drop in dynes cm. $^{-2}$; and $F_2(R_e, \sigma')$ is a function of R_e , the Reynolds number, and $\sigma' = \nu D$. In the expression for the heat transfer k , ν and ν_c , and here D and T , may with sufficient accuracy be taken as constants in any process which does not involve too large a difference of temperature or pressure. σ' is the analogue of the Prandtl number when the transfer of water vapour is considered instead of that of heat. Solution of the steady-state equations for the transfer of heat and water vapour to an unventilated sphere by conduction and diffusion respectively shows that for $R_e \rightarrow 0$,

$$F_1(R_e, \sigma) = F_2(R_e, \sigma') = 1.$$

Frössling (1938) has studied evaporation from spheres both theoretically and experimentally, and has shown that in the region $2 < R_e < 800$,

$$F_2(R_e, \sigma') = 1 + 0.276 \sqrt[3]{\sigma'} \sqrt{R_e}.$$

Sutton (1942) has confirmed this result. It will be seen later that the exact form of $F_2(R_e, \sigma')$ is not important, and this formula will be assumed valid down to $R_e = 0$. The largest cloud drops, $r = 100$ microns, have Reynolds numbers less than 10. With cloud drops the temperature and concentration differences are small. Under these conditions the equations governing the transfer of heat and water vapour, when expressed non-dimensionally, are identical, except that where σ' occurs in the case of diffusion, σ is substituted in the case of heat transfer. The solution of these equations applies equally to either, and the functions F_1 and F_2 must therefore be identical. We may therefore omit the suffixes and write

$$F(R_e, \sigma) = 1 + 0.276 \sqrt[3]{\sigma} \sqrt{R_e} = 1 + 0.246 \sqrt{R_e},$$

$$F(R_e, \sigma') = 1 + 0.276 \sqrt[3]{\sigma'} \sqrt{R_e} = 1 + 0.232 \sqrt{R_e}.$$

Howell (1949), quoting Langmuir (1944), has given a formula for the way in which the laws of diffusion and heat conduction change as the droplet radius becomes comparable with the mean free path of air and vapour molecules. It will be shown later that the variation of k and D with r may be neglected for, in general, the time taken by a droplet to grow beyond the size where this effect becomes small is short compared with that required for the formation of cloud drops of normal size.

Langmuir (1944) has shown that, with drops at rest with respect to the air, the separation of droplets in normal clouds is so great compared with their diameters that any interaction between them as regards the diffusion of water vapour or the conduction of heat may be neglected. In other words, the drops may be regarded as interacting independently with an atmosphere which has the same properties in the environs of each drop. When the motion of the drops is taken into account, this simplification is seen to be entirely justifiable. Imagine a cloud of average-sized drops of radius, say, 7 microns. They are separated on the average by more than 1000 microns, and the air moves through them at a speed of some 5000 microns per second. Obviously each drop will experience very nearly the same "average atmosphere", irrespective of its neighbours. Even very small drops of radius 1 micron fall at a speed of 100 microns per second. Any perturbations of the temperature or water-vapour concentration field due to other drops will be very small and transient. For the same reason, it does not seem necessary to discuss whether the flow of water vapour and heat about a drop departs from the steady state.

(b) *Equilibrium Vapour Pressure over a Drop*

The formula for the equilibrium vapour pressure over a drop according to the theory due to Köhler (1926) is

$$e_d = e(\theta_d) \left(1 + \frac{2\tau_w \varepsilon}{R \rho' r T} - \frac{3 M m_w}{4 \pi \rho' r^3} \right), \quad \dots \dots \dots (2.2)$$

where $e(0_d)$ is the equilibrium vapour pressure of a plane water surface at temperature 0_d , τ_w is the surface tension of water, ρ' its density, and m_w its molecular weight, and M is a measure of the size of the nucleus, which is assumed to be soluble. The positive term in the bracket expresses the increase in the vapour pressure of the drop due to capillarity; the negative one expresses the lowering due to the nucleus.

This formula is derived on the assumption that the solution is very dilute. On this basis, M is sometimes defined as the molar mass of the nucleus, the vapour-pressure lowering being found from Raoult's law. The small contribution of the solute to the total molar mass is neglected. Some authors point out that as the material of the nucleus (NaCl , H_2SO_4 , etc.) is an electrolyte, the molar mass must be multiplied by the number of ions formed when a molecule dissociates, that is, the van't Hoff factor is taken as being constant and equal to its value at infinite dilution. The value of τ_w is taken as that for pure water, in bulk. In this connection Tolman (1949) has discussed the variation of surface tension with drop size and concluded that, while it decreases with drop radius, the lowering amounts to as much as 4 per cent. only when the diameter is of the order of 10^{-6} cm. Kirkwood and Buff (1949) confirm this result.

The theory of drop growth will be based on formula (2.2). The assumptions mentioned and the definition of M will be reviewed later.

It is well known that the formula (2.2) leads to a maximum or critical value for e_d for a certain value of r . With slowly increasing vapour pressure, the nucleus drop grows steadily until the critical size is reached. In this stage of growth the nucleus is said to be unactivated and the drop is called a haze drop. Thereafter e_d decreases as r increases and the drop grows in an unstable fashion; this is the activated stage.

Howell (1949) has collated the measurements of several observers concerning the molar masses of nuclei. The smallest molar mass reported was $2 \cdot 10^{-22}$ and the largest 10^{-11} , the commonest size being 10^{-16} to 10^{-17} . Junge (1935) investigated the spectrum of condensation nuclei in surface air by means of graduated expansions, and concluded that most of them were activated by a supersaturation of 10 per cent. It will be seen that on the accepted theory this implies that most M values exceeded about 10^{-20} .

Wright (1939) has shown that sea-salt nuclei are considerably larger than combustion nuclei, and it is reasonable to suppose that the largest sea-salt nuclei would be found over the sea. Woodcock and Gifford (1949) and Woodcock (1950) have measured the masses of such nuclei by a deposition method. They found nuclei ranging up to about 10^{-9} g. from the sea surface up to cloud levels, the concentration decreasing with increasing height and nucleus size. For NaCl this corresponds to a maximum M value of rather less than 10^{-10} .

It will be assumed here that nuclei range in size from $M = 10^{-10}$ to $M = 10^{-22}$ and that cloud drops do not capture additional nuclei, that is, that M remains constant for a given drop. Langmuir (1948, Table 4) gives values for the collection efficiency of a larger drop overtaking a smaller one. This shows the collection efficiency increasing with the size of both cloud drops. However,

even drops of 100 micron radius collect no drops smaller than 3 micron radius. As coalescence of cloud drops with each other will be disregarded, the possibility of drops capturing additional nuclei may be neglected also.

(c) *Heat Balance of a Drop*

Condensation on the drop will release L calories per gram where L is the latent heat of condensation of water; this must be either lost by conduction or used to heat the drop. Sutton (1942) has considered the variation of temperature within a liquid drop. He found that the mean temperature of a water drop differs from the surface temperature by about $40r^2 \frac{d\theta_d}{dt} ^\circ\text{C}$. Now it will be seen that $d\theta_d/dt$ must approximate $d\theta/dt$, which itself must be nearly equal to $\Gamma'v$, where Γ' is the appropriate wet adiabatic lapse rate—say about $-6 \times 10^{-5} ^\circ\text{C. cm.}^{-1}$ —and v is the upward velocity of the air. Hence the difference must be about $0.002 vr^2$, which is evidently extremely small even for $r=100\mu$ and v large. Further, it will be seen later that another term of the same form as this one, but several times greater, is negligible.

The vapour pressure of the air e is equal to $e(\theta_s)$, the saturation vapour pressure at the dew-point θ_s . Using the Clausius-Clapeyron relationship a small vapour-pressure difference may be expressed in terms of a temperature difference :

$$e(\theta_s) - e(\theta_d) = e \varepsilon J L (\theta_s - \theta_d) / RT^2, \dots\dots\dots (2.3)$$

where J is the mechanical equivalent of heat, and e the average vapour pressure. Strictly e , L , and T should be taken at some temperature between θ_s and θ_d , but they may be taken with sufficient accuracy as being constant, provided too large a temperature range is not considered.

The heat balance of the drop is therefore expressed by

$$\frac{4}{3}\pi r^3 \rho' c \frac{d\theta_d}{dt} + 4\pi k F(R_e, \sigma) r (\theta_d - \theta) = \frac{4\pi \varepsilon^2 L^2 D J e}{R^2 T^3} F(R_e, \sigma') r \left(\theta_s - \theta_d - \frac{2\tau_w T}{J L \rho' r} + \frac{3 M m_w R T^2}{4\pi J L \varepsilon \rho' r^3} \right), \dots (2.4)$$

where c is the specific heat of water. It should be noted that this equation does not express the conservation of energy for the drop considered in isolation. It is for this reason that no term of the form $c\theta_d(dm/dt)$ appears.

The first term in (2.4) is small. To show this it may be written

$$\frac{d\theta_d}{dt} = (a + b\theta'_s) - (a + b)\theta_d, \dots\dots\dots (2.5)$$

where $a = 3kF(R_e, \sigma)/r^2 \rho' c$,

$$b = 3\varepsilon^2 L^2 D J e F(R_e, \sigma') / R^2 T^3 r^2 \rho' c,$$

$$\text{and } \theta'_s = \theta_s - \frac{2\tau_w T}{J L \rho' r} + \frac{3 M m_w R T^2}{4\pi J L \varepsilon \rho' r^3},$$

which expresses the effective dew-point of the air with respect to the drop.

Equation (2.5) may be solved in terms of integrals:

$$\theta_d - \frac{a\theta + b\theta'_s}{a+b} = \exp\left[-\int_{t_0}^t (a+b)dt\right] \left(\theta_d - \frac{a\theta + b\theta'_s}{a+b}\right)_{t=t_0} \\ - \exp\left[-\int_{t_0}^t (a+b)dt\right] \int_{t_0}^t \exp\left[\int_{t_0}^t (a+b)dt\right] \frac{d}{dt} \left(\frac{a\theta + b\theta'_s}{a+b}\right) dt.$$

Putting the functions F equal to unity in a and b gives their minimum values:

$$a = 1.79 \times 10^{-4} r^{-2}, \quad b = 3.36 \times 10^{-4} r^{-2}.$$

Hence the effect of any initial disturbance, expressed in the above solution by the first term, decreases with a time constant of about a quarter of a second even when $r = 100\mu$, and much faster for smaller drops. After a few seconds this term is quite negligible. Thus the value of θ_d is equal to its quasi-static value $(a\theta + b\theta'_s)/(a+b)$ minus another term which is evidently in the nature of a lag. The order of magnitude of this lag is most easily found from (2.5), since $d\theta_d/dt$ must approximate to $d\theta/dt$ which itself must be about equal to $\Gamma'v$. From (2.5)

$$\theta_d = \frac{a\theta + b\theta'_s}{a+b} - \frac{1}{a+b} \frac{d\theta_d}{dt},$$

and the lag term is equal to $\Gamma'v/(a+b)$ approximately, that is, about $0.14 \text{ } r r^2$. This is evidently extremely small for normal cloud drops.

Howell (1949), making step-by-step calculations of droplet growth, concluded that the thermal capacity of the drop could be neglected.

(d) Equation of Growth

Omitting the first term in (2.4), θ_d can be found in terms of θ and θ_s , and hence $\theta_s - \theta_d$. This gives the rate at which the drop grows as

$$\frac{dm}{dt} = 4\pi EF(R_s, \sigma') r \left(S - \frac{\beta}{r} + \frac{\gamma M}{r^3} \right), \dots\dots\dots (2.6)$$

where m is the mass of the drop,

$$E = \frac{\varepsilon^2 L D J e}{R^2 T^3} \left[1 + \frac{\varepsilon^2 L^2 D J e}{k R^2 T^3} \frac{F(R_s, \sigma')}{F(R_s, \sigma)} \right],$$

$$S = \theta_s - \theta,$$

$$\beta = 2\tau_w T / J L \rho',$$

$$\gamma = 3m_w K T^2 / 4\pi J L \varepsilon \rho'.$$

If for the lag of θ_d we write δ , the omission of the first term in (2.4) is equivalent to omitting a term $\frac{(a+b)}{a} \delta$ from the bracket on the right-hand side of (2.6), a and b having the same meaning as above, so that $(a+b)/a \approx 2.9$. Now $|\delta|$ is small except when $|r|$ is great, and then, as will be seen later, $|S|$ is large. The lag term will therefore be neglected (see Section IV (b)).

The ratio $F(R_s, \sigma')/F(R_s, \sigma)$ occurring in the denominator of E is equal to $(1 + 0.232 R_c^3)/(1 + 0.246 R_c^3)$, which is unity for R_c zero, and 0.97 for $R_c = 10$,

which is beyond the range of R_e with which we are concerned. E will therefore be regarded as a constant.

$S = \theta_s - \theta$ expresses the supersaturation of the air with respect to a plane water surface in terms of a dew-point difference. $S = 0$ corresponds to saturation. At a temperature of 10°C ., $S = 1^\circ\text{C}$. corresponds to about $6\frac{1}{2}$ per cent. supersaturation. The term $\left(S \cdot \frac{\beta}{r} + \frac{\gamma M}{r^3}\right)$ expresses the supersaturation of the air

with respect to a droplet of radius r , containing a dissolved nucleus of size M , the droplet being at the temperature of the air. In fact the droplet is warmer than the air when condensation is occurring. The physical meaning of this discrepancy is most easily seen when the terms β/r and $\gamma M/r^3$ are small compared with S , that is, where the drop is so large and the air is so far removed from saturation (in either direction) that the effects of capillarity and of the dissolved nucleus are negligible. In these circumstances the drop will take up the wet-bulb temperature of the air—apart from the very small lag mentioned above. At 10°C ., 800 mb., this will be rather more than the mean of θ and θ_s when $\theta_s > \theta$. The supersaturation available to cause condensation on the drop is only about one-third of that of the air. Combining the expression (2.1) and equation (2.3) shows that this is taken into account by the denominator of E , which is in fact approximately equal to 3. Obviously the same applies, *mutatis mutandis*, when evaporation is occurring. It is therefore essential to take account of the heat balance of the drop in computing its growth or evaporation.

On the other hand, as is implied in neglecting the first term of (2.4), the heat content of the drop is quite negligible compared with the quantities of heat released by condensation, and conducted away by the air, during any significant growth. It has been suggested that cloud drops can grow—even into raindrops—by condensation, because they are initially cooler than the environment, as a result of mixing, or because they are falling into successively warmer layers. Both these ideas appear to be based on a misconception of the relative orders of magnitude of the heat content of a cloud droplet and the heat exchange which goes on as a result of condensation and conduction to the air. Suppose, for example, that as a result of turbulent mixing, a droplet moves into a saturated environment $\frac{1}{2}\theta^\circ\text{C}$. warmer than itself. It will warm up as a result of both condensation and conduction, and these two effects will in fact be of about the same magnitude. However, leaving aside the conduction effect, we have

$$Ldm = mc d\theta_a$$

during the condensation process. Hence when the drop has reached temperature equilibrium its relative increase in mass will be

$$\frac{\Delta m}{m} = e^{(c/L)\Delta\theta} - 1 \approx \frac{c}{L}\Delta\theta.$$

Thus for $\frac{1}{2}\theta = 1^\circ\text{C}$. the proportional increase in mass is 1.6 per cent., in radius 1.18 per cent. This is of course an over-estimate, since the drops would in fact be heated by conduction also. This is entirely negligible compared with the growth which is necessary to form large cloud drops or raindrops. Unless a

mechanism can be found by which the droplet can be re-cooled without evaporation occurring in the converse manner, and the process repeated over and over, no appreciable growth can occur in this way. Similarly, if a cold drop falls through successively warmer layers and is heated altogether by, say, 30 °C., even if we attribute all this heating to the release of latent heat of condensation, and neglect conduction from the air, the total increase in mass cannot exceed about 5 per cent. The parallel which is sometimes drawn between the supposed growth of an initially cold droplet in this manner and the growth of ice crystals in a supercooled water cloud is misleading. The ice crystals will take up a temperature very close to the ice-bulb temperature of the air, which will be less than the frost point, but greater than the temperature of the air, since the latter is supersaturated with respect to ice. In this way the ice crystals can dispose of the latent heat of sublimation by conduction without losing the property which causes the sublimation to occur. As in the case of water droplets in supersaturated air, however, this warming above the air temperature reduces the supersaturation available with respect to the growing particle which, in both cases, is an essential prerequisite for significant growth.

The effect of ventilation on the growth of a drop is expressed by the factor $F(R_e, \sigma')$ in (2.6). Dady (1950) has shown that eddies contribute very little to the ventilation of a cloud drop in comparison with gravity. Thus R_e is a function of drop size and (to a much smaller extent) of temperature and pressure. It is given as a function of the drop radius in Table 1, for stated conditions; it differs little for other conditions. The values of R_e are computed for freely falling spheres of unit density at the terminal velocity. Even at $r = 500\mu$, the velocity so computed agrees fairly well with measured values such as those of Schmidt (1909) and Sutton (1942).

TABLE 1
VALUES OF THE VENTILATION FACTOR $F(R_e, \sigma')$ IN (2.6) FOR 10 °C., 800 MR.

$r(\mu)$	10	30	50	70	100
R_e	0.014	0.37	1.5	3.2	8.0
$F(R_e, \sigma')$	1.03	1.14	1.28	1.41	1.66
$(\sqrt{F(R_e, \sigma')} - 1)$	0.01	0.07	0.13	0.19	0.29

The factor does not differ significantly from unity for normal cloud drops. If the growth of a drop up to a radius of, say, 50 μ is computed, neglecting ventilation, the time so found for the growth to occur will be in error by less than 28 per cent. In the case where the capillarity and nucleus terms in (2.6) are negligible, the proportional error in r cannot exceed $(\sqrt{F(R_e, \sigma')} - 1)$ shown in the last line of the table.

In deriving equation (2.6) the heat balance of the drop and the effect of ventilation have been taken into account; temperature differences within the drop and changes in the heat content of the drop have been disregarded as being negligible. However, (2.6) describes only the growth of an individual

cloud drop in air of known supersaturation. In order to trace the history of the drops in a parcel of cloud air, it is necessary to determine how the supersaturation itself changes.

III. THE GROWTH OF DROPLETS IN A CLOUD

(a) *The Adiabatic Assumption*

The variation of the supersaturation, S , can be determined in a simple way only if we assume, in effect, that a portion of the cloud may be isolated, in a closed box as it were, with no transfer of heat or matter between it and the environment. Then the conservation of energy and mass determine the variation of S in terms of changes in height and in liquid water present. This assumption is equivalent to neglecting the falling velocities of the droplets with respect to the air, and is reasonable only if the "box" is envisaged as large compared with the total displacement of the largest drops relative to the air during the period considered. On the other hand, the "box" must not be imagined as so big that turbulence in the cloud would, during the period considered, destroy its integrity. The extent to which these calculations are reasonable therefore depends on whether dimensions can be chosen for the "box" which fulfil both these conditions, which in turn depends on the time involved in the process considered, *inter alia*.

If w is the mass of liquid water associated with a gram of dry air, and x the humidity mixing ratio, the conservation of water is expressed by

$$dx + dw = 0.$$

Transforming dx into terms of de and dp , and thence into terms of $d\theta_s$ and dz , and using the energy relationship, we find (neglecting the small energy contents of the water vapour and liquid water)

$$dS = q_1 dz - q_2 dw, \dots\dots\dots (3.1)$$

where

$$q_1 = \frac{g}{J} \left(\frac{1}{c_p} - \frac{T}{L\varepsilon} \right),$$

$$q_2 = \frac{L}{c_p} + \frac{pRT^2}{JL\varepsilon^2 e}.$$

It is evident that the coefficients of the differentials here may be regarded as constant, provided the process considered does not involve too great a vertical movement. Similarly in (2.6) above, the only variables which need be considered are m , t , R_e , S , r , and M .

(b) *Equations of Droplet Growth in a Cloud*

Substituting numerical values corresponding to 10 °C. and 800 mb. for the remaining parameters, the equations (2.6) and (3.1) become

$$dm/dt = 8.28 \times 10^{-7} (1 + 0.232 R_e^{1/2}) r (S - 1.719 \times 10^{-6} r^{-1} + 64.36 M r^{-3}),$$

$$dS = 7.930 \times 10^{-5} dz - 4.016 \times 10^3 dw,$$

where $dw = \sum dm$ summed over the droplets associated with 1 gram of dry air.

The equations (2.6) and (3.1) or their numerical forms given above determine completely the evolution of a droplet spectrum with any nucleus distribution and any velocity under the adiabatic assumption. They are capable of numerical treatment only. In the present state of knowledge of the spectra of condensation nuclei and of motions in cloud, any numerical specification of the initial conditions—chief among which is the nucleus spectrum—and of the subsequent motion, would be very arbitrary. Any attempt to reach general conclusions in this way—beyond those found by Kraus and Smith (1949) and Howell (1949)—is unpromising. It seems desirable therefore to study any general characteristics of the condensation process which can be demonstrated without unduly specializing the conditions.

IV. SOME GENERAL CHARACTERISTICS OF CONDENSATION PROCESSES IN CLOUD

(a) *An Approximate Value for the Supersaturation*

A general conclusion can be drawn about the values assumed by S without integrating the equations completely, as follows: addition of the equation (2.6) for all the drops (in one gram of air) gives

$$dw/dt = 4\pi E(S\chi_1 - \beta\chi_0 + \gamma\chi_{-2}), \dots\dots\dots (4.1)$$

where

$$\chi_1 = \Sigma rF(R_e, \sigma'),$$

$$\chi_0 = \Sigma F(R_e, \sigma'),$$

$$\chi_{-2} = \Sigma Mr^{-2}F(R_e, \sigma').$$

Combining this with (3.1) gives

$$\begin{aligned} dS/dz = q_1 - 4\pi E q_2(S\chi_1 - \beta\chi_0 + \gamma\chi_{-2})/v, \\ = a_1 - a_2 S, \dots\dots\dots (4.2) \end{aligned}$$

where

$$a_1 = q_1 + 4\pi E q_2(\beta\chi_0 - \gamma\chi_{-2})/v,$$

$$a_2 = 4\pi E q_2 \chi_1/v.$$

If a_1 and a_2 are regarded as functions of z , this may be solved to give S as a function of z , in terms of integrals involving a_1 and a_2 . The solution may be transformed to give the following result:

$$S = S_0 e^{-a_2(t-t_0)} + \left(\frac{a_1}{a_2}\right)_t (1 - e^{-a_2(t-t_0)}),$$

where t_0 is an arbitrary initial instant, $t_0 < t < t$, and

$$a_2 = 4\pi E q_2 \int_{t_0}^t \chi_1 dt / (t - t_0) = 4\pi E q_2 \chi_1.$$

To find S , t_0 may be chosen arbitrarily, but the corresponding value for S_0 must be inserted. If, however, $a_2(t - t_0)$ is large enough, the exponential becomes small so that the effect of S_0 becomes negligible and S may be taken as $(a_1/a_2)_t$. This formula will therefore yield useful results when a_2 is large enough. Then $(t - t_0)$ need not be chosen so great that $(a_1/a_2)_t$ tends to be too indefinite.

Now $4\pi E q_2 = 3.325 \times 10^{-3}$, so that the effect of S_0 decreases with a time constant of $301/\chi_1$. But

$$\chi_1 = \Sigma r F(R_e, \sigma') > \Sigma r,$$

which in normal clouds is of the order of 50 to 500 cm. g.⁻¹ (counting only the observable drops). Hence the time constant is of order 1 to 10 seconds. Taking $\chi_1 = 50$, the exponential decreases to 0.1 in 14 seconds; for $\chi_1 = 500$, in 1.4 seconds. Equation (4.2) shows that the error in equating S to a_1/a_2 is equal to

$$\frac{1}{4\pi E q_2 \chi_1} \frac{dS}{dt} = \frac{301}{\chi_1} \frac{dS}{dt},$$

that is, approximately equal to the change in S during an interval of one time constant. If dS/dt is small, a_1/a_2 affords a good approximation for S . Now

$$\frac{a_1}{a_2} = \frac{q_1}{4\pi E q_2 \chi_1} v + \frac{\beta \chi_0}{\chi_1} - \frac{\gamma \chi^{-2}}{\chi_1}.$$

The χ summations extend over all drops, however small. But if drops smaller than a certain limiting size (r_1) are neglected, (4.1) shows that the

resulting error is $\frac{1}{4\pi E \chi_1} \frac{dw_1}{dt}$, where w_1 is the mass of liquid water in the form of such drops. If we assume that growth or evaporation among these small drops is equivalent, at the most, to the formation from zero size of n_1 drops per gram of radius r_1 per second, the error in a_1/a_2 is $n_1 r_1^3 / 3E \chi_1$; for $r_1 = 0.5 \mu$, $n_1 = 50,000$ drops per gram per second, $\chi_1 > 50$ cm. g.⁻¹, this is less than 10^{-3} °C. (0.006 per cent. R.H.). This may be regarded as negligible, for it will be seen later (Figs. 4 and 5) that this error would in general have no significant effect on drop growth.

The first term in (a_1/a_2) expresses the supersaturation necessary to maintain condensation at the rate $(q_1 v / q_2)$ g. g.⁻¹ sec.⁻¹, neglecting the effect of capillarity and of the nucleus. This rate of condensation would hold S itself constant (see (3.1)). The second term expresses the additional supersaturation required to overcome the effect of capillarity in order to maintain the same rate of condensation. The third term gives the reduction in the necessary value of S due to the nuclei, which of course assist condensation. The first term is equal to $0.0239v/\chi_1$. The second is nearly equal to $\beta n/\chi_1$, where n is the number of drops per gram; this is equivalent to an increment of v of $7.2 \times 10^{-5} n$. Taking $n = 3 \times 10^5$ g.⁻¹ (≈ 300 drops per cc.) this is equivalent to some 22 cm. sec.⁻¹. The effect of this term will therefore be appreciable, relatively, when v is small. But if \bar{r} is the arithmetic mean of the radii, it is nearly equal to $\beta/\bar{r} = 1.72 \times 10^{-6}/\bar{r}$. In normal clouds, therefore, its absolute magnitude is small (for $\bar{r} > 5 \mu$, less than 3.5×10^{-3} °C.). The third term is nearly equal to $\beta \Sigma (r_s/r)^2 / \chi_1$, where $r_s = \sqrt{\gamma M / \beta}$, the equilibrium drop radius at $S=0$, or

$\frac{\beta}{3\chi_1} \Sigma (r_c/r)^2$, where r_c is the critical radius ($\sqrt{3\gamma M / \beta}$). This term is therefore $\frac{1}{3} \frac{\Sigma (r_c/r)^2}{n}$ times the second. If the cloud drops are well beyond the activation

stage, so that $r \gg r_c$ for all the drops, this ratio will be small. In that case we find for S

$$S \approx (0.0239v + 1.72 \times 10^{-6}n)/\chi_1. \quad \dots\dots\dots (4.3)$$

Of course $\chi_1 \approx \Sigma r$, provided there are not too many large drops present. The formula (4.3) applies in the body of the cloud where S is not changing too quickly, and where all, or almost all, the drops are well beyond their critical stage. It is of particular interest that under these conditions, and provided drops which are too small to be observed are responsible for only a small proportion of the total water economy, S can be related to quantities which are all measurable, the approximate ranges of which, in nature, are known. This is independent of any special assumptions regarding velocity or (with a reservation to be made later) of the nucleus spectrum. The statement has often been repeated that the supersaturation in cloud cannot reach values large enough to be measurable by means of a hair hygrometer or similar instrument, on the grounds that no such supersaturation has ever been detected. But it is difficult to see what justification can exist for this view. Any measuring element exposed to cloud air with an appreciable relative velocity will become wet in a few seconds owing to the impact of drops. Thereafter, it can hardly be expected to indicate other than saturation. High supersaturations are most likely to occur in strong upcurrents; in such cases it would be very difficult in practice to keep the instrument at rest with respect to the air. Taking values which could occur in vigorous convection: $v = 1000$ cm. sec.⁻¹, $r = 5 \cdot 10^{-4}$ cm., $n = 3 \cdot 10^5$ drops per gram, we find from (4.3): $S = (23.9 + 0.5) 150 = 0.16$ °C. (1 per cent. R.H.). When $v \gg 7.2 \cdot 10^{-5}n$ (say, more than a few metres a second) the first term in the numerator of (4.3) is dominant, and

$$S \approx 0.0239v/\chi_1. \quad \dots\dots\dots (4.4)$$

Evidently, except for very small velocities when the capillarity term $\beta\chi_0/\chi_1$ becomes important, S is larger (for a given water content) when there are a few large, rather than many small, drops.

The simple formula (4.3) for S applies only when $\Sigma(r_c/r)^2 \ll n$, which may not always be the case. Thus

$$r_c^2/r^2 = 3\gamma M/\beta r^2 = 1.12 \times 10^8 M/r^2.$$

Provided $M = 10^{-18}$, $r = 0.5 \cdot 10^{-4}$ cm., this will be small compared with unity. Of course, large values of M will tend to be associated with large values of r , but, in the absence of any knowledge of the nuclei occurring in a group of observed cloud drops, it must be realized that the third term might become important. Suppose for example that the n drops formed around equal nuclei of size M . Then the ratio of the third term to the second would be $0.37 \cdot 10^8 M/r^2$. If we assume that these larger nuclei will have formed drops of $r = 10\mu$ in the body of the cloud, this term becomes comparable to the second when $M = 10^{-14}$; for $M = 10^{-12}$, its neglect would cause an error in S of 0.064 °C. (0.4 per cent. R.H.). Of course it is unlikely that any but a very small proportion of the nuclei would be as large as this, and the error in S would be correspondingly smaller.

Obviously no direct experimental check can be made on these formulae. We may, however, use the calculations of Howell (1949), who gives several sets of curves showing the history of drop sizes and supersaturation in uniformly cooled air. These all agree in showing that the supersaturation rises to a maximum near the cloud base and then decreases steadily.

The quasi-static value for S , namely (a_1/a_2) , and the approximations (4.3) and (4.4) have been evaluated for Howell's three computations for his conditions (0°C. , 800 mb.). The results are compared with the supersaturation curves given by him in Figure 1. The dashed curve (a_1/a_2) agrees quite well with the value given by Howell after S has passed its maximum. It will be noted that the curve (a_1/a_2) should pass through the turning point of S .

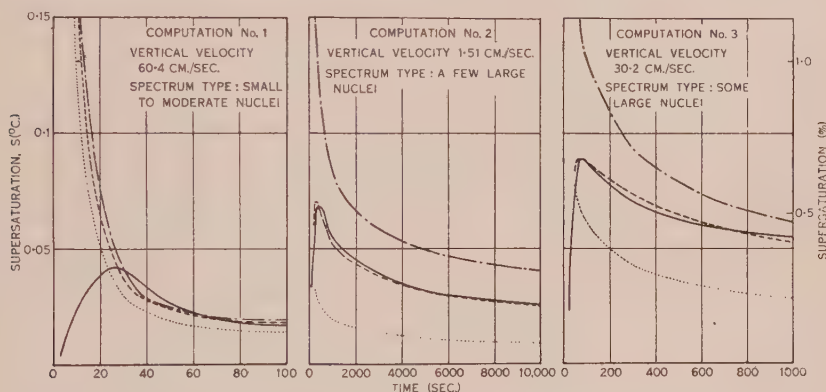


Fig. 1.—Comparison of approximate formulae for the supersaturation (S) with Howell's curves.

- Howell's supersaturation curve.
- a_1/a_2 (the quasi-static value).
- · - · - The value from (4.3) (neglecting the nuclei).
- The value from (4.4) (neglecting the nuclei and capillarity).

since dS/dz vanishes there. Thereafter it should lie below the S curve, since S is decreasing. The lack of exact correspondence in these respects may perhaps be attributed to slight differences in the values of physical constants. The approximations (4.3) and (4.4) are satisfactory in the first case where there are no nuclei larger than $M=10^{-15}$ and the velocity is fairly high. In the second case where there are a few large nuclei and the velocity is very small, and in the third case where the velocity is moderate, but there are more large nuclei, these curves are seriously in error.

(b) The Relative Magnitude of the Lag Term Neglected in (2.6)

The approximate value now found for S may be compared with the lag term neglected in deriving (2.6). This is approximately $2.9\delta = 0.41vr^2$. For $r < 100\mu$, $|v| < 80$ cm. sec. $^{-1}$, allowing for ventilation $|2.9\delta| < 2 \times 10^{-3}^\circ\text{C.}$, which is equivalent to an error of the same amount in S , and may be regarded as negligible, as was pointed out above. For $v > 80$ cm. sec. $^{-1}$, the simple formula (4.4) will be a fair approximation to S in the body of the cloud. The

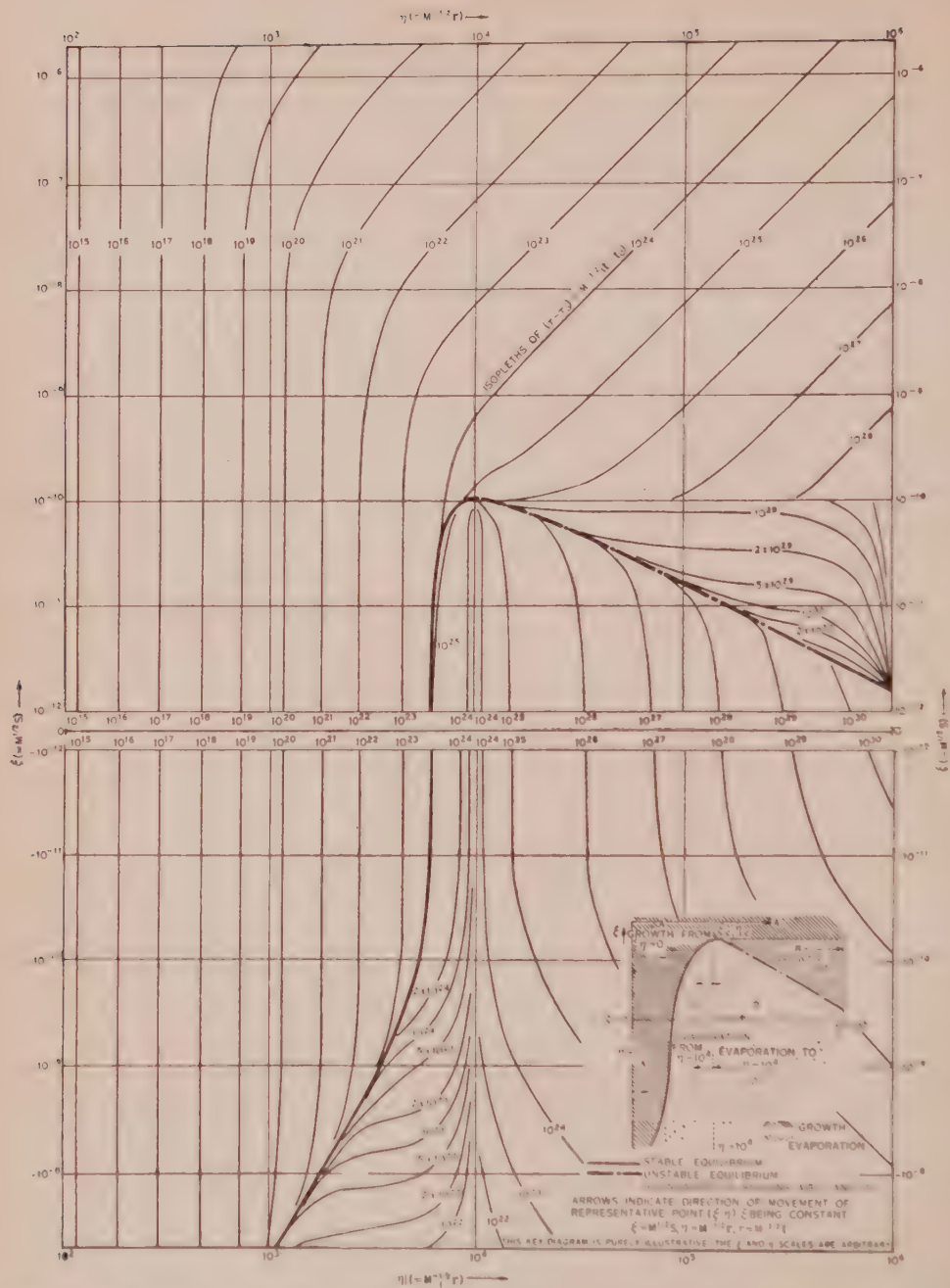


Fig. 2. The growth or evaporation of unventilated drops at constant supersaturation— isopleths of $(\tau - \tau_0)$ from equation (4.6).

ratio $2.9\delta S$ is therefore approximately equal to $17r^2/\lambda_1$. Large drops will be associated with small values of λ_1 . The order of magnitude of the error may be found in terms of w , by assuming the drops are all equal. Putting $\lambda_1 = nr$,

$$2.9\delta/S \approx 4w,$$

which corresponds to a maximum of about 2 per cent. for $w = 5 \times 10^{-3}$ (≈ 5 g. m. $^{-3}$). Thus the neglect of the lag term is equivalent to an error of the order of a few per cent. in S at most. If S dominates in (2.6) the same error will appear in dm/dt . If not, the error will, in general, be smaller still.

(c) Integration of the Equation of Growth with Constant Supersaturation

Some general characteristics of cloud droplet growth may be derived by integrating (2.6) with S constant and $F(R_e, \sigma') = 1$. The magnitude of the errors introduced by neglecting ventilation can be seen from Table 1. The results of this integration do not of course apply directly to the growth of droplets in a cloud, where S will vary, but they can be used to find orders of magnitude, to set limits, and to review in a quantitative way some of the assumptions made in deriving (2.6).

It is advantageous to introduce new variables defined by

$$\xi = M^{1/2} S, \quad \eta = M^{-1/2} r, \quad \tau = M^{-3/2} t.$$

Equation (2.6) then becomes

$$\frac{\eta^4 d\eta}{\xi \eta^3 - \zeta \eta^2 - \gamma} = \frac{E}{\xi} d\tau, \quad \dots \dots \dots (4.5)$$

in which M no longer appears explicitly, so that a single set of solution curves can be used whatever the value of M . The solution is given by

$$\begin{aligned} \frac{E\xi}{\rho}(\tau - \tau_0) = & \frac{1}{2}(\eta^2 - \eta_0^2) + \frac{\beta}{\xi}(\eta - \eta_0) + A \log \frac{\eta - \lambda_1}{\eta_0 - \lambda_1} \\ & + B \log \frac{\eta - \lambda_2}{\eta_0 - \lambda_2} + C \log \frac{\eta - \lambda_3}{\eta_0 - \lambda_3}, \quad \dots \dots \dots (4.6) \end{aligned}$$

where λ_1, λ_2 , and λ_3 are the roots (real or complex) of the cubic

$$\xi \eta^3 - \zeta \eta^2 - \gamma = 0,$$

and

$$A = \lambda_1^4(\lambda_2 - \lambda_3) \triangleq, \quad B = \lambda_2^4(\lambda_3 - \lambda_1) \triangleq, \quad C = \lambda_3^4(\lambda_1 - \lambda_2) \triangleq,$$

where

$$\triangleq = -\lambda_1^2(\lambda_2 - \lambda_3) - \lambda_2^2(\lambda_3 - \lambda_1) - \lambda_3^2(\lambda_1 - \lambda_2).$$

The solution curves giving $(\tau - \tau_0)$ as a function of η are given in Figure 2 for the conditions where the temperature is 10°C . and the pressure 800 mb.

By means of these curves the time can be found for growth or evaporation at constant supersaturation (S) between any two sizes, for any size of nucleus. Drop radius is measured by the abscissae ($r = M^{1/2}\eta$), supersaturation by the ordinates ($S = M^{-1/2}\xi$), and time by the isopleths ($(t - t_0) = M^{3/2}(\tau - \tau_0)$).

The heavy curve corresponds to the equilibrium condition. It corresponds to the family of curves usually drawn to show the variation with radius of the

equilibrium vapour pressure over a drop for various sizes of nuclei (see, for example, Köhler 1926; Wright 1936; Howell 1949). The stable portion of the curve is shown solid, the unstable, broken.

As the supersaturation is taken as constant, ξ remains constant during any growth, and the representative point of a drop moves horizontally on the figure as t and τ increase. The direction of movement is shown by arrows in the inset, being to the right for points lying in the shaded regions *A* and *B* above the heavy equilibrium curve (growth) and to the left for points in the stippled regions *C* and *D* below (evaporation).

In the region *A* shown on the inset, the times given are for growth from zero size ($r=0$, $\eta=0$). The region *B*, however, cannot be reached from zero size with constant supersaturation. Here, the times shown are for growth to an arbitrary size ($\eta=10^6$). Below the heavy equilibrium curve, drops evaporate. As, however, the drop radius ($-\eta M^{1/2}$) reaches the stable equilibrium value, or departs finitely from the unstable curve only after an infinite time these curves cannot be used as starting points. The line $\eta=10^4$ has been chosen as a reference; the regions each side of it are labelled *C* and *D*. The line $\eta=10^4$ passes practically through the critical point at the apex of the equilibrium curve (critical point coordinates: $\xi_c=2\beta^{3/2}\gamma^{-1/2} \approx 3 \times 10^{-10} \times 10^{-10}$; $\eta_c = \sqrt{3}\gamma^{-1/2} \approx 1.06 \times 10^4$). In region *C*, the times shown are for evaporation from $\eta=10^4$; in *D* for evaporation to $\eta=10^4$, as indicated in the inset.

The growth or evaporation time of a drop from radius r_1 to r_2 is therefore computed as follows. The corresponding values of η ($\eta_1 = M^{-1/2}r_1$, $\eta_2 = M^{-1/2}r_2$) and the value of ξ for the ambient supersaturation ($= M^{1/2}S$) are found. Let the corresponding values of $(\tau - \tau_0)$ taken from the figure be called τ_1 and τ_2 . Then the time taken for a growth (regions *A* or *B*) is $M^{3/2}|\tau_2 - \tau_1|$ and for evaporation (from region *D* to region *C*), $M^{3/2}(\tau_2 + \tau_1)$.

The region *A* of Figure 2 is the simplest and also the most directly interesting. The properties of droplet growth in this region are seen more easily from the curves of Figure 3. These show how radius ($r = M^{1/2}\eta$) increases with time ($t = M^{3/2}\tau$) for a particular constant supersaturation ($S = M^{-1/2}\xi$). These growth curves differ fundamentally according as $\xi > \xi_c$. For $\xi > \xi_c$, $\eta \rightarrow \infty$ with τ . Otherwise η tends to its equilibrium value for the particular value of ξ . In the former case the growth curves asymptote to the lines

$$\eta^2 = 2E\xi\tau/\rho', \quad \dots\dots\dots (4.7)$$

which is the solution of (4.5) when the term $\xi\eta^3$ dominates in the denominator; that is when $|\xi|$ and η are large. Thus for $|\xi|$ and η large enough, capillarity and the nucleus exert no influence. This corresponds to air far removed from saturation (in either direction), large drop size, and a small nucleus.

The maximum value of ξ shown (10^{-5}) corresponds to $M=10^{-10}$, $S=1$, which are about the maximum values we need consider. When η is small and ξ not too great, the $\xi\eta^3$ term in the denominator of (4.5) becomes negligible, and the solution is independent of ξ , and nearly equal to that for $\xi=0$:

$$\frac{\beta E \tau}{\rho'} = \eta_s^3 \left\{ \frac{1}{5} \left(\frac{\eta}{\eta_s} \right)^5 + \frac{1}{7} \left(\frac{\eta}{\eta_s} \right)^7 + \dots \right\}.$$

Figure 3 shows that for $0 < \xi < 10^{-6}$ and $\eta < 3 \times 10^2$, the growth curves do indeed run very close together. Further, for $\eta < 3 \times 10^2$, the above solution lies within 1 per cent. of the curve

$$\frac{\beta E \tau}{\rho'} = \frac{1}{5} \frac{\eta^5}{\eta_s^2},$$

or

$$\eta^5 = 5E\gamma\tau/\rho', \quad \dots\dots\dots (4.8)$$

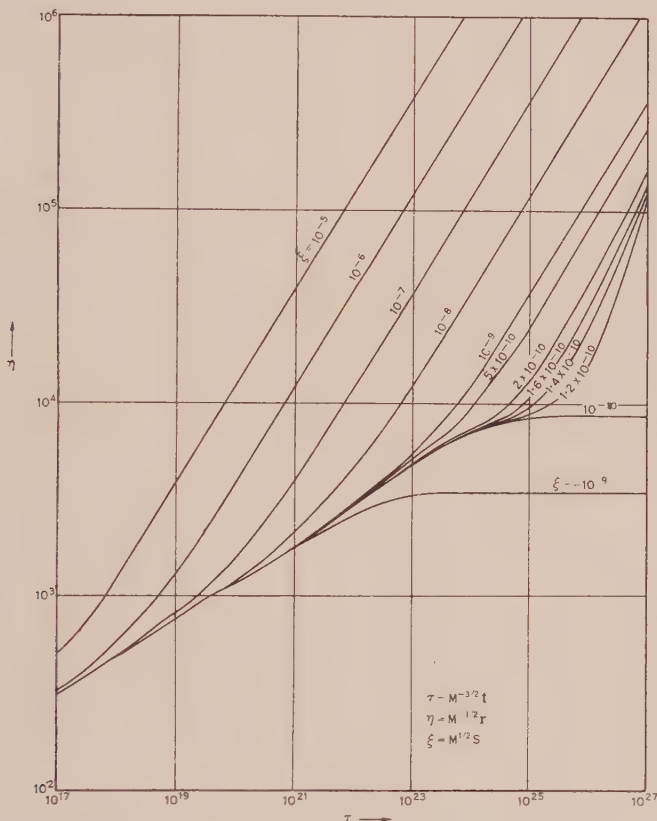


Fig. 3.—Growth curves at constant supersaturation, from zero size (from Fig. 2).

which is the solution of (4.5) neglecting $\xi\eta^3$ and $\beta\eta$. This is the solution when the nucleus dominates the drop's growth. It corresponds to air close to saturation, small drop size, and a large nucleus. By means of the simple formulae (4.7) and (4.8), the growth of the drops outside the limits of Figure 2 can be computed without difficulty. These results will now be used to deduce some general characteristics of cloud droplet growth.

(d) Droplet Growth at Small Sizes

Imagine a haze drop suddenly subjected to some positive supersaturation. No matter how small it may have been initially, at the end of one second it will

have reached at least the size indicated by Figure 2 for growth from $\eta=0$. Table 2 shows the result.

In general therefore, after one second a drop either approaches very closely its equilibrium size, or else grows to a radius at least of the order of 1μ . This generalization is of importance in the following respect. In the computation

TABLE 2

GROWTH OF DROPS FROM ZERO SIZE DURING ONE SECOND FOR POSITIVE SUPERSATURATIONS. (ACCORDING TO FIG. 2, AND EQUATION (4.7). $S_c = M^{-1/2} \epsilon_c$ IS THE CRITICAL SUPERSATURATION)

M	τ	Growth of Drop
10^{-22}	10^{33}	Practically in equilibrium if $S < S_c$; $r > 12\mu$ if $S > 1.1 S_c$
10^{-20}	10^{30}	Practically in equilibrium if $S < S_c$; $r > 3.6\mu$ if $S > 1.1 S_c$
10^{-18}	10^{27}	Practically in equilibrium if $S < S_c$; $r > 0.9\mu$ if $S > 1.1 S_c$
10^{-16}	10^{24}	Practically in equilibrium if $S < S_c$; $r > 0.7\mu$ if $S > 1.1 S_c$
10^{-14}	10^{21}	$r > 1.8\mu$ in all cases
10^{-12}	10^{18}	$r > 4.6\mu$ in all cases
10^{-10}	10^{15}	$r > 12\mu$ in all cases

leading to Figure 2, D and k were treated as independent of r . In fact the "effective" values of D and k decrease with r below sizes of the order of 1μ . This does not affect the equilibrium size of a drop, which occurs when heat conduction and vapour diffusion have ceased: $c_d = c_w$, $\theta_d = \theta$. It does, however, change the rate at which a drop grows. Table 2 suggests that the error caused by treating D and k as constant would be unimportant, for, above $r = 1\mu$, the effective values of D and k change comparatively little.

Howell (1949) gives formulae for "compensated" coefficients of diffusion and heat conduction (D' and k'). These give $D' = D$, $k' = k$ for $r \gg \lambda_w$, λ_a (the mean free path of air and water-vapour molecules). For $r \ll \lambda_w$, λ_a , D' and k' are proportional to r , so that the rate of diffusion and conduction to a drop is proportional to r^2 , that is, to its surface area. The following are the equivalents of Howell's formulae:

$$D' = \left[\frac{r}{D(r + \alpha\lambda_w)} + \frac{1}{r} \sqrt{\frac{2\pi\epsilon}{RT}} \right]^{-1},$$

$$k' = \left[\frac{r}{k(r + \alpha\lambda_a)} + \frac{4}{r\rho c_p f} \sqrt{\frac{\epsilon}{3RT}} \right]^{-1},$$

where α is the Cunningham constant (0.7), f is the accommodation coefficient of air molecules colliding with liquid water (0.7).

With D' and k' dependent on r , E (equation (2.6)) is no longer constant, and will be written E' . The values of D' , k' , E' , and $E/E' (= E'(\infty)/E'(r))$ are given in Table 3 for 10°C ., 800 mb.

The calculations of time required for a specified growth may be corrected to allow for the variation of D and k by putting

$$\delta t' = \frac{E}{E'} \delta t,$$

where δt refers to the time computed from Figure 2, and $\delta t'$ refers to the corrected time for growth from r to $r + \delta r$. As $E' \rightarrow 0$ as $r \rightarrow 0$, it is convenient to begin the

TABLE 3

COMPENSATED COEFFICIENTS OF DIFFUSION (D') AND HEAT CONDUCTION (k') ABOUT A SPHERE OF RADIUS r AT 10 °C., 800 MB., AND THE RESULTING VALUE OF E' (EQUATION (2.6))

r (cm.)	10^{-6}	2×10^{-6}	5×10^{-6}	10^{-5}	2×10^{-5}	5×10^{-5}	10^{-4}	2×10^{-4}	5×10^{-4}	∞
$D' \cdot 10^{-2} \times$	1.43	2.81	6.49	11.1	16.5	22.9	26.1	28.0	29.2	30.1
$k' \cdot 10^{-6} \times$	2.59	5.13	12.1	21.1	32.5	45.6	52.0	55.7	58.1	59.7
$E' \cdot 10^{-9} \times$	2.95	5.82	13.6	23.6	35.9	50.2	57.2	61.4	63.9	65.9
E/E'	22.4	11.3	4.84	2.79	1.83	1.31	1.15	1.07	1.03	1.00

calculations at some non-zero radius; $r=10^{-6}$ cm. has been chosen. Nuclei are usually bigger than this in normal air, and in any case, calculations based on a constant surface tension τ_w , and a constant value of the van't Hoff factor would not be significant below such sizes. Proceeding by finite steps gives an upper limit for the time taken for growth from r to $r + \delta r$, since E/E' is a monotonic decreasing function of r . Calculations carried out in this manner yield the curves of Figure 4.

For various values of S , the corrected and uncorrected growth times (t and t') are shown for different values of M in the range 10^{-12} to 10^{-18} . The initial value of r is taken as 10^{-6} cm. $M=10^{-20}$ does not appear, for even at $S=1$ such nuclei do not grow beyond $r=10^{-6}$ cm. It is obvious that the time differences, measured horizontally between the t and t' curves, are never physically significant in considering the growth to cloud drop size. In general, therefore, the curves of Figure 2, based on the use of constant values of D and k , may be used for computations of the growth of cloud drops. The differences between the t and t' curves are important, *relatively*, only below $r=1\mu$. The only case where significant differences arise for growth beyond 1μ is when S is very slightly greater than S_c , and $r_c \ll 1\mu$ ($M \ll 10^{-16}$). Then the drop may spend many seconds growing through sizes less than 1μ , where the factor E/E' is considerably greater than unity. This can happen, however, only for a small range of S values (see Table 2) and a slight change in S would cause the drop either to grow quickly beyond 1μ or else to tend rapidly towards an equilibrium size.

The characteristics of drop growth with $S > 0$ at small sizes which were pointed out from Table 2 remain valid; drops tending to equilibrium sizes approach their final value very closely in a few seconds, while those which are growing in an unstable manner pass beyond $r=1\mu$ in a period of the order of one second, however small S may be. The latter characteristic is surprising. It is explained by the fact that, with S small, only large nuclei can become

unstable, and their influence on c_d persists up to sizes of the order of $r=1\mu$, and larger. In Figure 4, most of the curves for $S=10^{-3}$, 10^{-4} reflect the fifth power law of (4.8) which applies when the nucleus is dominant.

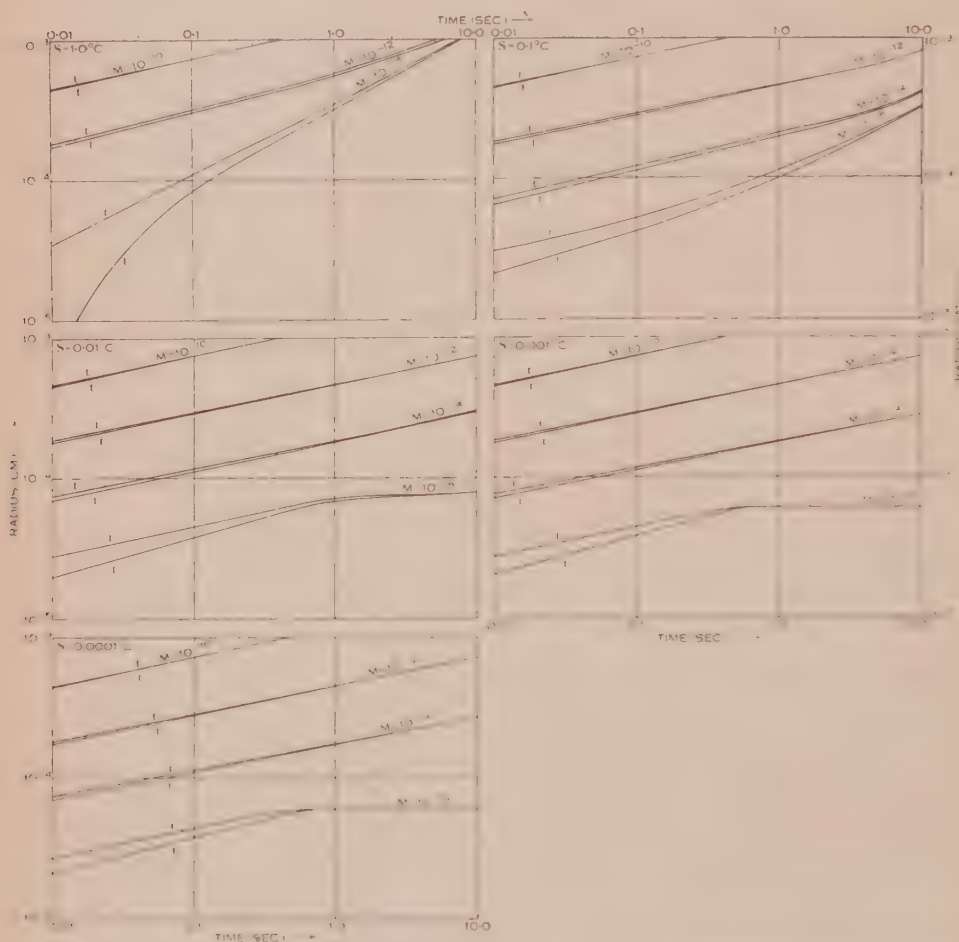


Fig. 4.—Growth curves from $r=10^{-6}$ cm. for various values of the supersaturation (S) and nucleus size (M): t —time as found from Figure 2 (constant E); t' —corrected times taking account of the variation of E with r . Cases such as $M=10^{-18}$, $S=10^{-3}$ are not shown, for the drop comes practically into equilibrium at a radius smaller than 10^{-5} cm.

(e) Droplet Growth to Large Sizes

In Section II (c) it was shown that an essential prerequisite for significant growth is that the air should be supersaturated with respect to the growing drop. The presence of a nucleus ensures that this will be so even in saturated air when the drop is small enough, and the question naturally arises whether very large nuclei have the power of forming large cloud drops, or even raindrops, without the assistance of supersaturation (with respect to a plane water surface).

If the air is just saturated, $S=0$. The equilibrium drop radius is then

$$r_s = M^{\frac{1}{3}} \eta_s = 6 \cdot 12 \times 10^3 M^{\frac{1}{3}}.$$

For $M=10^{-12}$, $r_s=61 \cdot 2\mu$, which corresponds to a large cloud drop, and for $M=10^{-10}$, $r_s=612\mu$, which corresponds to raindrop size common in light rain. This size is not, however, attained in a finite time. The growth of droplets and very large nuclei up to $r=100\mu$ for $S=0$ is shown in Table 4.

TABLE 4
GROWTH FROM ZERO SIZE OF CLOUD DROPLETS AROUND VERY LARGE
NUCLEI AT SATURATION ($S=0$) (NO VENTILATION)

r (μ)	Time Taken (seconds)		
	M		
	10^{-12}	10^{-11}	10^{-10}
10	50	3	0.5
30	16,000	1,250	122
50	200,000	20,000	5,100
70	—	73,200	10,000
100	—	537,000	50,000

Ventilation would have the effect of reducing the period of time required, but not by as much as 40 per cent. even for the biggest drops. Very long periods are required for even the biggest nuclei to form drizzle drops without the assistance of some degree of supersaturation.

The effect of varying degrees of supersaturation on nuclei of various sizes is shown in Figure 5 after the lapse of 100, 1000, 3000, and 10,000 seconds. These curves show how the final value of $r(r_f)$ depends on M for various values of S . It is immediately evident that very large nuclei ($M=10^{-10}$, 10^{-11}) can form large cloud drops in moderately long time-intervals with little or no supersaturation present. Smaller nuclei require the assistance of an appreciable degree of supersaturation. The most striking feature of these curves, however, is the sudden change in r_f near the critical value of M when S is large. For $S \leq 0$, there is no critical phenomenon and the curves are quite smooth. Where this threshold effect is marked, practically all the activated nuclei grow to much the same final size. This corresponds, of course, to the results quoted in the Introduction. Suppose that drops of a certain final radius—say $r_f=30\mu$ —are to be found. If a period of 3000 seconds is available, and nuclei as large as $M=10^{-11}$ are present, no supersaturation is necessary; indeed a subsaturation of $0 \cdot 01^\circ \text{C}$. suffices. Under these circumstances r_f varies steadily with M . But if the growth is to occur within 1000 seconds, and the largest nucleus present is $M=10^{-12}$, a supersaturation of $0 \cdot 07^\circ \text{C}$. is essential. In this case, practically all the activated nuclei form drops of radius 25μ or larger. Thus, for a given

final radius, the importance of the threshold effect increases as the growth time and the size of the largest nucleus decreases.

If large cloud drops are to be formed at all, it is essential that there should not be too many of them to share the liquid water released. Taking air saturated at 20 °C. at the 900 mb. level, with a lifting of 10,000 feet, if $r_f = 30\mu$, the number

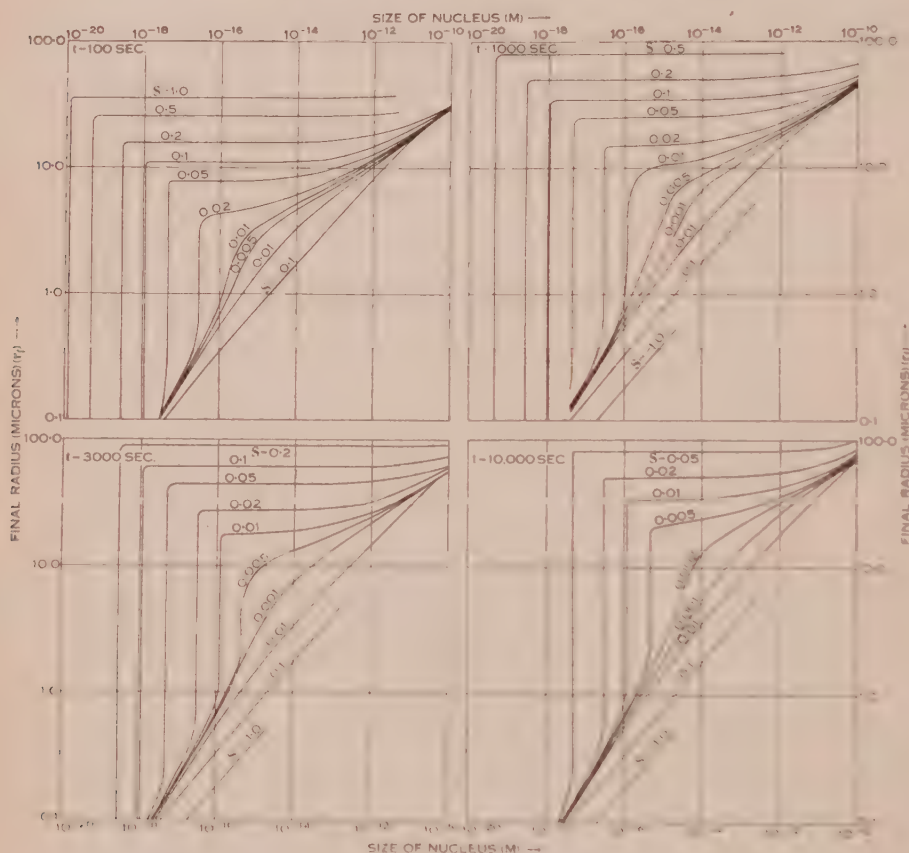


Fig. 5.—The effect of the nucleus size (M) on growth at constant supersaturation (S) to large drop sizes. Curves showing the final radius (r_f) reached after various intervals at various supersaturations as a function of the nucleus size.

of large drops cannot exceed about 60 per cc. The types of nucleus spectrum suitable for the formation of large drops will be discussed in more detail in Part II of this series.

The effect of ventilation, which becomes significant for the larger drop sizes, has been neglected in computing the curves of Figure 5. This would displace the curves upwards, but would not alter the general appearance of the figure; the threshold effect in particular would still appear. In their quasi-horizontal portions the curves would undergo a change of shape, but this would be second order.

(f) *The Effect of Temperature*

The values of η_c , ξ_c , and E are given for various temperatures in Table 5.

It is clear that the critical supersaturation of a nucleus ($M^{-\frac{1}{3}}\xi_c$), its critical radius ($M^{\frac{1}{3}}\eta_c$), and its equilibrium radius at saturation ($M^{\frac{1}{3}}\eta_c/\sqrt{3}$) all vary only slightly with the temperature. On the other hand, E nearly doubles itself between 0 and 30 °C. For a given effective supersaturation, expressed in °C., the rate of mass growth of a drop increases quite strongly as the temperature increases. The effect of this increase on the growth of drops in cloud will be discussed in Part II of this series.

TABLE 5
EFFECT OF TEMPERATURE ON THE CONSTANTS η_c , ξ_c , AND E

Constant	Temperature (°C.)			
	0	10	20	30
$\eta_c \dots 10^4 \times \dots$	1.053	1.060	1.089	1.119
$\xi_c \dots 10^{-10} \times \dots$	1.051	1.081	1.080	1.078
$E \dots 10^{-8} \times \dots$	4.906	6.590	8.028	9.185

It will be noted that while the critical supersaturation of a nucleus expressed in °C. varies little with temperature, the corresponding supersaturation expressed as relative humidity increases by a factor of about 2.5 between 0 and 30 °C.

(g) *The Definition of M, the Measure of Nucleus Size for Small Nuclei*

The usual theory, which has been followed here, computes e_d , the vapour pressure of a drop, on the assumption that it consists of a very dilute solution. Then the van't Hoff factor (i) is constant and the vapour-pressure depression due to the nucleus is proportional to the molar concentration. It is easy to see that this is not always true even for drops in equilibrium with saturated air. The number of moles of nucleus (or of effective moles after ionization) being defined as M , the theory given above leads to an equilibrium droplet radius of $\eta_c M^{\frac{1}{3}}$ so that the number of moles of nucleus per litre of solution must be

$$\frac{3 \times 10^3 M^{-\frac{1}{3}}}{4\pi\eta_c^3} \approx 10^{-9} M^{-\frac{1}{3}}.$$

For very small values of M like 10^{-20} this is no longer so small that i is constant. Indeed, for $M=10^{-20}$, this would represent a supersaturated solution in the case of NaCl. Again, Table 4 shows how slowly a very large nucleus approaches equilibrium. At its equilibrium size at $S=0$, a drop formed on a nucleus $M=10^{-12}$ has a radius of 61.2μ , with a nuclear concentration of 10^{-3} moles per litre of solution. At $r=10\mu$, the concentration would therefore be 0.23 mole per litre. The values of i for various concentrations are given in Figure 6

(computed from data in International Critical Tables and Landolt-Börnstein) for sodium chloride and sulphuric acid. The variation in i is equivalent to a variation in M with drop size. This variation in the M value of an individual nucleus is of course entirely negligible compared with the total range of possible M values.

An additional complication arises from the variation of surface tension with concentration. In the case of NaCl for example, this amounts to about 1.5

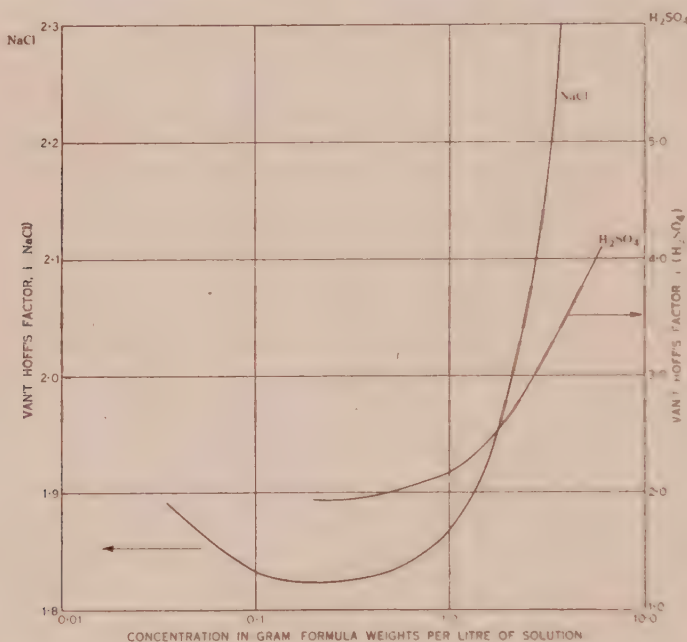


Fig. 6.—The variation of van't Hoff's factor, i , with concentration.

dynes cm.⁻¹ per mole per litre (Kaye and Laby 1944). The critical supersaturations of very small nuclei are slightly increased as a result.

For these reasons the theory is only approximate, especially for small nuclei. Now the most important and most easily observed property of a small nucleus is its critical supersaturation. It would seem therefore, that it is the best one to use as a measure of the size of small nuclei. If this is done, M may be defined in terms of an observed critical supersaturation, S_c :

$$M = S_c^2 / \xi_c^2,$$

where ξ_c is computed as before. The theoretical nucleus for which i and τ_c are taken as constant will then have the correct value of S_c , but its critical radius and its behaviour at other supersaturations will differ from those of the actual nucleus.

Defining M as suggested, the simplicity of the usual theory may be retained at the cost of some theoretical accuracy. There seems little doubt that it describes the main physical effects at cloud drop sizes with an accuracy adequate

to our present knowledge of condensation nuclei. Any more precise theory would involve the chemical constitution of the nucleus. In the present state of knowledge, such an additional complication would be of little use.

V. CONCLUSION

The equation of growth of an individual drop has been derived taking into account the effect of ventilation due to its falling velocity. Owing to the motion of the drops through the air, interaction between the drops, and possible non-stationary conditions in the flow of heat and water vapour, may be disregarded. The effect of ventilation is to increase the rate of growth, but it does not become serious until the drops are fairly large (Table 1).

The heat economy of a growing cloud drop is completely dominated by the balance between the heat released by condensation and that conducted away from the drop. Changes in its heat content are relatively negligible, and no appreciable growth can occur merely because a drop is falling through successively warmer layers of saturated air, or because mixing brings it into contact with warmer air. On the contrary, an essential condition is that it should be warmer than the air. Otherwise the heat released by condensation cannot be disposed of. For this reason, only a certain fraction of the actual supersaturation of the air with respect to a drop (at the same temperature) is available to cause condensation. At 10 °C., 800 mb., this is about one-third.

The equations governing the evolution of a cloud droplet spectrum are constructed under the assumption that the vertical motions occur in an adiabatic manner. The subsequent discussion aims at clarifying some general characteristics of condensation processes in cloud which do not depend on special assumptions about the vertical velocity or the nucleus spectrum.

Near the condensation level, the phenomena in cloud are complex and depend critically on the nucleus spectrum. In the body of a cloud, where most of the growing drops have already passed their critical size, the situation is simpler. Here the supersaturation tends towards its quasi-static value (i.e. the value it would have to have if it were constant) with a time constant of the order of one to ten seconds. Therefore, except where it is changing rapidly, the quasi-static value provides a good approximation. This value can, in general, be expressed approximately in terms of observable quantities—the number of drops, their radii, and the vertical velocity. Typical values of these yield supersaturations ranging up to a few per cent. at most. The method fails when many very large nuclei are present.

In order to discuss the characteristics of cloud droplet growth, the equation of growth is integrated for constant supersaturation, neglecting ventilation. The results are applied to droplet growth at small and large sizes. It is found that, at small sizes, drops either approach their equilibrium size quite closely, or else grow to a size of the order $r=1\mu$ in a period of a few seconds. For this reason, it is justifiable to neglect the variation of the laws of diffusion and heat conduction at diameters of the order of the mean free path of air molecules,

for the normal laws apply from radii of about 1μ upwards, without serious error. Drops can grow to sizes large compared with normal cloud drops in the following circumstances :

- (i) At small supersaturations : there must be a few very large nuclei present, and growth must continue for several thousand seconds.
- (ii) At moderate supersaturations : all but a few of the nuclei must then be so small as not to become activated, for all activated nuclei grow to much the same size. Periods of the order of one thousand seconds and even less are sufficient.

The definition of the size of a small nucleus is reviewed. It is shown that even in saturated air the solution in a droplet is not always very dilute. Consequently van't Hoff's factor is not by any means constant during the growth of all drops, although it is usually assumed to have a constant value corresponding to infinite dilution. To retain the simplicity of the usual theory it is suggested that the most satisfactory definition of the size of a small nucleus would be based on the value of the critical supersaturation.

VI. LIST OF SYMBOLS

A, B, C , see equation (4.6),

D , diffusivity of water vapour in air,

D' , effective diffusivity about a small sphere,

E , see equation (2.6),

E' , see Table 3 (effective value of E for very small drops),

$F(R_s, \sigma)$ } ventilation factors for transfer of heat and water vapour to spheres.
 $F(R_s, \sigma')$ }

J , mechanical equivalent of heat,

L , latent heat of condensation of water vapour,

M , the measure of the size of a condensation nucleus (of the order of molar mass of nucleus),

R_s , Reynolds number,

R , gas constant of dry air, per gram,

S , supersaturation in $^{\circ}\text{C.} = \text{dew-point } (\theta_s) - \text{temperature } (\theta) \text{ of air,}$

S_c , critical supersaturation of a nucleus,

T , absolute temperature (average),

V , falling velocity of a droplet,

a, b , see equation (2.5),

c , specific heat of water,

c_p , specific heat of air at constant pressure,

e , water-vapour pressure (average),

e_a , water-vapour pressure of air,

e_d , equilibrium vapour pressure over a drop,

f , accommodation coefficient of air molecules colliding with liquid water (0.7),

g , acceleration due to gravity,

- i , van't Hoff's factor (see Sections II (b), IV (g)),
 k , thermal conductivity of air,
 k' , effective conductivity about a small sphere,
 m , mass of a droplet,
 m_w , molecular weight of water,
 n , number of drops per gram of dry air,
 p , total pressure of air and water vapour,
 q_1, q_2 , see equation (3.1),
 r , radius of a droplet,
 r_c , critical radius of a drop,
 r_f , final radius of a drop (Fig. 5),
 r_s , equilibrium radius in saturated air,
 t , time,
 v , upward velocity of air,
 w , liquid water content in grams per gram of dry air,
 z , height,
 Γ' , wet adiabatic lapse rate,
 Δ , see explanation of equation (4.6),
 $\chi_0, \chi_1, \chi_{-2}$, see equation (4.1),
 α , Cunningham constant (0.7),
 β , see equation (2.6),
 γ , see equation (2.6),
 δ , the lag of θ_d behind its quasi-static value (negligible),
 ε , specific gravity of water vapour with respect to dry air,
 η , see equation (4.5) ($=M^{-\frac{1}{2}}r$),
 η_s , equilibrium value of η in saturated air,
 η_c , critical value of η ($=\sqrt{3}\beta^{-\frac{1}{2}}\gamma^{\frac{1}{2}}=1.06\times 10^4$),
 θ , temperature of air,
 θ_s , dew-point of air,
 θ_d , temperature of a drop (strictly, of the surface),
 $\lambda_1, \lambda_2, \lambda_3$, see equation (4.6),
 λ_w, λ_a , mean free path of water vapour and air molecules respectively,
 ν , kinematic viscosity of air,
 ν_c , thermal diffusivity of air,
 ρ , density of air,
 ρ' , density of liquid water,
 σ , Prandtl number,
 σ' , analogue of σ , referring to diffusion of water vapour,
 τ , see equation (4.5) ($=M^{-3/2}t$),
 τ_w , surface tension of water,
 ξ , see equation (4.5) ($=M^{\frac{1}{2}}S$),
 ξ_c , critical value of ξ ($=2\beta^{3/2}\gamma^{-\frac{1}{2}}/3\sqrt{3}=1.081\times 10^{-10}$).

VII. REFERENCES

- BYERS, H. R., and BRAHAM, R. R. (1948).—*J. Met.* **5** : 71-86.
- DADY, G. (1950).—*J. Sci. Met.* 2^e Année No. 8 : 108-20.
- FİNDEISEN, W. (1939).—*Met. Z.* **56** : 453-60.
- FRÖSSLING, N. (1938).—*Beitr. Geophys.* **52** : 170-216.
- GOLDSTEIN, S. (1938).—“Modern Developments in Fluid Dynamics.” 1st Ed. Vol. 2. p. 607. (Clarendon Press : Oxford.)
- HOWELL, W. E. (1949).—*J. Met.* **6** : 134-49.
- JUNGE, C. (1935).—*Met. Z.* **52** : 467-70.
- KAYE, G. W. C., and LABY, T. H. (1944).—“Tables of Physical and Chemical Constants.” 9th Ed. p. 46. (Longmans, Green and Co.: London.)
- KIRKWOOD, J. G., and BUFF, F. P. (1949).—*J. Chem. Phys.* **17** : 338-43.
- KÖHLER, H. (1926).—*Medd. Met.-Hydr. Anst. Uppsala* **3** : No. 8.
- KRAUS, E. B., and SMITH, BETTY (1949).—*Aust. J. Sci. Res. A* **2** : 376-88.
- LANGMUIR, I. (1944). Supercooled water droplets in rising currents of cold saturated air. Gen. Elect. Res. Lab. P 2984 (6), Precipitation Static Studies, Oct. 1943-Aug. 1944.
- LANGMUIR, I. (1948).—*J. Met.* **5** : 175-92.
- SCHMIDT, W. (1909).—*Met. Z.* **26** : 183-4.
- SÜRING, R. (1927).—“Leitfaden der Meteorologie.” p. 125. (Leipzig.)
- SUTTON, O. G. (1942). Investigations on falling drops carried out at the Chemical Defence Experimental Station, Porton. Air Minist. Met. Res. Cttee. MRP No. 40.
- TOLMAN, R. C. (1949).—*J. Chem. Phys.* **17** : 333-47.
- WOODCOCK, A. H., and GIFFORD, MARY M. (1949).—*J. Mar. Res.* **8** : 177-97.
- WOODCOCK, A. H. (1950).—*J. Met.* **7** : 161-2.
- WRIGHT, H. L. (1936).—*Proc. Phys. Soc.* **48** : 675-89.
- WRIGHT, H. L. (1939).—*Quart. J.R. Met. Soc.* **65** : 411-42.

THE INFLUENCE OF METEOROLOGICAL CONDITIONS ON RAINMAKING IN THE SYDNEY AREA

By E. J. SMITH*

[*Manuscript received August 28, 1951*]

Summary

An analysis is presented of the results of 88 flights whose purpose was to carry out rainmaking experiments, using dry ice, in the Sydney area between January 1948 and June 1950. The chance of encountering clouds suitable for rainmaking is shown to be correlated with the synoptic situation, the water content above the freezing level, the height of the freezing level, and the lapse rate below it. The number of days on which rainmaking was possible in the Sydney area is estimated as 45, 33, and 8 for 1948, 1949, and January to June 1950 respectively, a smaller number of opportunities occurring when the natural rainfall was abnormally high. The possibility of forecasting days suitable for rainmaking is examined.

I. INTRODUCTION

Investigations into the artificial stimulation of precipitation by seeding clouds with dry ice have been carried out in Australia since early in 1947. The methods of experiment and observation and some of the results have previously been published (Kraus and Squires 1947 ; Smith 1949, 1950 ; Squires and Smith 1949).

It was found that precipitation could be released from cumulus or alto-cumulus cloud, consisting of supercooled water droplets, provided the cloud was sufficiently durable and cold. The process took 10 to 25 minutes, depending on the depth of cloud, and it was therefore necessary for the cloud to have a life in excess of that figure. Whenever success was achieved, the temperature of the clear air level with the cloud top was colder than -2°C . and usually -7°C . If the rain were to reach the ground, the cloud needed to be compact, with the top vertically above the bottom so that precipitation particles could grow as they fell through the cloud, and also deep, with the height of the base above ground less than two-thirds the height of the top.

A considerable number of experiments were carried out in the Sydney area, in which attempts were made to cause rain to fall where none would have fallen naturally. Observations were made of the types of cloud encountered, whether or not they were suitable for rainmaking. In the present paper an attempt is made to determine the meteorological factors which were associated with various types of cloud, and in particular with those suitable for rainmaking.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

II. DESCRIPTION AND CLASSIFICATION OF FLIGHTS

Between January 1, 1948 and June 30, 1950, flights have been made with the primary purpose of carrying out experiments in the artificial stimulation of precipitation. All those considered in this paper took place within 150 miles of Sydney, over land or a few miles out to sea.

A total of 88 flights was made, for which all the necessary data are available. In a number of other flights technical failures occurred or observations were inadequate: also some flights were made in other localities: all these have been discarded. The analysis is concerned with the number of flights on which various phenomena were observed: it is not concerned with the number of such phenomena observed per flight. For example, if on a flight two or more successful experiments were carried out, this is classed as "one successful flight".

The results of these flights have been classified into seven categories, as follows.

- A.* One or more successful experiments in which dry ice was dropped into a deep cumulus cloud which was not precipitating. Precipitation occurred within 25 minutes and reached the ground, and is believed to have been caused by the dry ice. There was no natural precipitation within 30 miles.
- B.* Successful experiment in which precipitation was similarly released from alto-cumulus cloud, but evaporated before reaching the ground.
- C.* Suitable cloud but unsuccessful experiment ("suitable" means apparently similar to those described above under *A* or *B*, with temperature in the clear air level in the top colder than -2°C). Dry ice was dropped but no results were observed.
- D.* Suitable cloud but no experiment performed.
- E.* All supercooled clouds encountered precipitated naturally, or were less than 30 miles from others which did.
- F.* All supercooled clouds encountered were transient: the part above the freezing level rose and rapidly evaporated or sank again.
- G.* No supercooled clouds encountered.

Of these, group *A* (successes where rain reached the ground) is of practical importance, while groups *A* to *D* inclusive are of experimental interest. Groups *E*, *F*, and *G* are not suitable for dry-ice rainmaking experiments.

Some flights appeared to fall into two categories—for example, there may have been supercooled clouds, some of which were transient, while others rained naturally. In these cases the flight was included only once, in the highest category in the list in which it fitted. If both alto-cumulus and deep cumulus were available for experiment, the cumulus was always selected.

In the absence of any proved method of forecasting the occurrence of suitable clouds, flights were made with only slight regard to the weather conditions. However, when there was no secondary objective, flights were sometimes cancelled if there were either general rain or no clouds in the experimental area, and when excessive local rain prevented the use of the aerodrome. Sampling of conditions was therefore almost random, with a tendency to omit the most unsuitable days.

TABLE 1
RESULTS AND METEOROLOGICAL CONDITIONS FOR EACH FLIGHT

Date	Time of Day	Result	Type of Synoptic Situation	Height of Freezing Level above Normal (ft.)	Water Content 2000 ft. above Freezing Level (g./kg.)	Difference in Height between 0 and +10 °C. Isotherms (ft.)
1948						
19. ii	p.m.	<i>C</i>	<i>G</i>	0	2.7	5200
10. iii	p.m.	<i>A</i>	<i>G</i>	-2300	3.3	4200
16. iii	p.m.	<i>C</i>	<i>G</i>	-4300	1.9	5300
18. iii	p.m.	<i>E</i>	<i>G</i>	-2000	0	4800
1. iv	p.m.	<i>C</i>	<i>G</i>	-1500	3.2	4800
7. iv	p.m.	<i>G</i>	<i>B</i>	+1100	0	7400
13. iv	p.m.	<i>A</i>	<i>G</i>	-5000	3.6	4400
14. iv	p.m.	<i>C</i>	<i>G</i>	-2000	2.8	3200
15. iv	p.m.	<i>A</i>	<i>G</i>	-4000	3.4	4300
27. iv	p.m.	<i>B</i>	<i>I</i>	-2000	3.3	5700
28. iv	p.m.	<i>F</i>	<i>I</i>	-1200	2.4	4500
1. vi	p.m.	<i>G</i>	<i>B</i>	-1000	1.6	6600
7. vi	a.m.	<i>G</i>	<i>B</i>	-1000	3.0	5200
9. vi	p.m.	<i>F</i>	<i>B</i>	-1000	3.0	4900
11. vi	p.m.	<i>B</i>	<i>G</i>	-900	3.4	4400
17. vi	a.m.	<i>B</i>	<i>B</i>	-200	4.2	6500
28. vi	a.m.	<i>G</i>	<i>B</i>	-3000	0	11500
3. viii	p.m.	<i>G</i>	<i>B</i>	+800	0	5900
5. viii	a.m.	<i>G</i>	<i>B</i>	0	1.4	7500
11. viii	p.m.	<i>E</i>	<i>B</i>	-1000	0	4700
17. viii	p.m.	<i>A</i>	<i>I</i>	-2000	3.3	3600
19. viii	a.m.	<i>F</i>	<i>B</i>	-500	2.5	3700
25. viii	a.m.	<i>A</i>	<i>G</i>	-2800	0.7	4000
31. viii	a.m.	<i>B</i>	<i>B</i>	-700	1.4	3800
15. ix	a.m.	<i>G</i>	<i>B</i>	-1000	1.3	5000
23. ix	a.m.	<i>E</i>	<i>G</i>	-200	3.7	5700
5. x	p.m.	<i>E</i>	<i>B</i>	-2500	3.8	5300
6. x	p.m.	<i>G</i>	<i>B</i>	-2000	1.7	3700
12. x	p.m.	<i>G</i>	<i>G</i>	-1000	1.8	6000
13. x	a.m.	<i>B</i>	<i>I</i>	-2500	2.7	5300
19. x	p.m.	<i>G</i>	<i>B</i>	+2000	0.8	10200
22. x	a.m.	<i>A</i>	<i>G</i>	-4600	2.2	4000
1. xi	p.m.	<i>C</i>	<i>B</i>	-3800	3.4	3700
3. xi	a.m.	<i>E</i>	<i>B</i>	-1200	1.0	5000
4. xi	p.m.	<i>F</i>	<i>B</i>	0	0	5700
25. xi	p.m.	<i>A</i>	<i>G</i>	-2200	4.0	3800
1. xii	a.m.	<i>G</i>	<i>B</i>	-800	0	8200
2. xii	p.m.	<i>A</i>	<i>G</i>	+700	2.0	5200
8. xii	a.m.	<i>E</i>	<i>B</i>	-1700	0.9	6300
14. xii	p.m.	<i>G</i>	<i>B</i>	-500	0	10500
15. xii	a.m.	<i>G</i>	<i>B</i>	+700	0	5400
1949						
26. i	a.m.	<i>G</i>	<i>B</i>	-2500	2.8	8300
27. i	a.m.	<i>G</i>	<i>B</i>	-2000	1.3	7000
4. ii	a.m.	<i>B</i>	<i>I</i>	-1600	3.3	4100

TABLE 1 (Continued)
RESULTS AND METEOROLOGICAL CONDITIONS FOR EACH FLIGHT

Date	Time of Day	Result	Type of Synoptic Situation	Height of Freezing Level above Normal (ft.)	Water Content 2000 ft. above Freezing Level (g./kg.)	Difference in Height between 0 and +10 °C. Isotherms (ft.)
1949						
9. ii	p.m.	<i>E</i>	<i>G</i>	-200	4.0	5000
17. ii	p.m.	<i>G</i>	<i>B</i>	-1000	1.2	6800
18. iii	p.m.	<i>E</i>	<i>I</i>	+1700	0	11800
22. iii	p.m.	<i>G</i>	<i>B</i>	-600	0	8800
24. iii	p.m.	<i>G</i>	<i>B</i>	-1300	1.4	7100
30. iii	a.m.	<i>A</i>	<i>G</i>	-4400	2.7	4900
14. iv	a.m.	<i>G</i>	<i>B</i>	-2000	1.3	4400
3. v	p.m.	<i>E</i>	<i>B</i>	-1000	2.8	4900
5. v	p.m.	<i>A</i>	<i>G</i>	-6200	2.0	4000
11.viii	p.m.	<i>B</i>	<i>G</i>	-500	0	4000
15.viii	p.m.	<i>D</i>	<i>G</i>	-2500	2.4	4000
30.viii	a.m.	<i>G</i>	<i>B</i>	+700	1.1	5900
1. ix	p.m.	<i>A</i>	<i>G</i>	-1800	2.9	3700
27. ix	a.m.	<i>G</i>	<i>B</i>	+700	0	8300
29. ix	a.m.	<i>E</i>	<i>B</i>	-1500	2.3	4700
11. x	p.m.	<i>G</i>	<i>B</i>	+2000	0	6500
18. x	p.m.	<i>E</i>	<i>B</i>	+300	0	6000
22. xi	p.m.	<i>E</i>	<i>B</i>	-1200	2.6	5000
24. xi	p.m.	<i>F</i>	<i>G</i>	-2000	1.9	7000
1. xii	a.m.	<i>G</i>	<i>I</i>	-3500	1.8	5800
13. xii	a.m.	<i>D</i>	<i>I</i>	-1500	2.5	4200
15. xii	a.m.	<i>G</i>	<i>B</i>	-600	0	10300
1950						
8. ii	a.m.	<i>E</i>	<i>B</i>	-1000	3.7	6400
13. ii	a.m.	<i>G</i>	<i>I</i>	-2300	1.1	6800
16. ii	a.m.	<i>G</i>	<i>I</i>	-200	1.5	6700
17. ii	a.m.	<i>G</i>	<i>B</i>	-1200	3.8	8300
21. ii	a.m.	<i>G</i>	<i>I</i>	+700	0	6600
23. ii	a.m.	<i>G</i>	<i>B</i>	+500	0	7000
1. iii	a.m.	<i>G</i>	<i>G</i>	+1200	0	6500
3. iii	a.m.	<i>E</i>	<i>I</i>	-800	3.5	4500
6. iii	a.m.	<i>F</i>	<i>G</i>	-200	1.2	5600
20. iii	p.m.	<i>E</i>	<i>B</i>	+2000	0	7800
24. iii	a.m.	<i>G</i>	<i>I</i>	+1000	0	7300
19. iv	a.m.	<i>F</i>	<i>I</i>	-1000	2.8	5800
21. iv	a.m.	<i>G</i>	<i>B</i>	-700	0	7300
26. iv	a.m.	<i>G</i>	<i>I</i>	+200	1.0	7300
1. v	a.m.	<i>E</i>	<i>B</i>	+1200	0	8500
3. v	a.m.	<i>E</i>	<i>B</i>	0	3.5	6200
4. v	a.m.	<i>D</i>	<i>I</i>	700	3.4	5000
8. v	a.m.	<i>E</i>	<i>B</i>	-1000	1.1	5800
16. v	a.m.	<i>E</i>	<i>B</i>	-1300	0	7000
27. v	a.m.	<i>E</i>	<i>G</i>	-300	1.6	6300
31. v	a.m.	<i>E</i>	<i>B</i>	+3000	0	7900
5. vi	p.m.	<i>A</i>	<i>I</i>	-1300	3.2	5700

The results of these flights, in categories as described above, are given in Table 1 (column 3). Columns 4 to 7 refer to the conditions under which flights were made. They are discussed in subsequent paragraphs.

III. EXAMINATION OF DATA

An attempt was made to find out which meteorological factors were favourable for rainmaking by a statistical analysis of the conditions when flights in the different categories were made. The following factors were examined:

- (a) Synoptic chart.
- (b) Height of freezing level.
- (c) Water content at heights from 6000 ft. above the freezing level to 8000 ft. below it.
- (d) Average lapse rate between $+10$ and 0°C .
- (e) Average lapse rate between 0 and -10°C .
- (f) Wind shear between 1500 and 10,000 ft.
- (g) Season.
- (h) Time of day.

The data for (b), (c), (d), and (e) were obtained from the radiosonde soundings made daily at 1800 hours E.A.S.T. from Rathmines, 70 miles NNW. of Sydney.

The accuracy of many of the observations was not very high, and of course preparation and examination of synoptic charts are subjective processes. With these reservations it appeared that the observed conditions were correlated with (a), (b), (c), and (d), in a manner described in the following paragraphs, but no correlation could be observed with the other factors.

The wind shear was examined by comparing the wind at about 1500 and 10,000 ft. measured respectively from the isobars of the surface and the contours of the 700 mb. charts. No correlation was found with the clouds which were observed. However, as mentioned in Section I, clouds whose tops are not vertically above their bottoms have proved to be unsuitable for rainmaking. It therefore seems that wind shear may be important, but the data examined were not sufficiently accurate or comprehensive to reveal this.

There was approximately the same number of flights in the mornings and afternoons, but more flights in the afternoons than the mornings fell in category *A* (see Section II) when successful experiments were performed on deep cumulus cloud, and also in the combined categories $A+B+C+D$ which are experimentally useful. However, if flights are considered separately on days which are suitable or unsuitable for rainmaking as measured by the "rainmaking index" described in Section VIII, there was no difference in the incidence of conditions suitable for rainmaking in the mornings and the afternoons. It is not possible to be sure without further flights whether the time of day influenced the conditions observed.

There was no significant difference in the incidence of suitable conditions for rainmaking in the different seasons. It may be that season and time of day are interrelated, for example, summer afternoons and winter mornings might be favourable. The flights were too few and too unevenly distributed to enable this point to be investigated.

IV. SYNOPTIC SITUATIONS NEAR SYDNEY

(a) The Typical Situation

A characteristic feature of Australian weather is a series of anticyclones which cross the continent from west to east. These have been described by Kidson (1925). On the average they are separated by about 2000 miles and move at about 300 miles per day. In the summer they usually pass to the south of Sydney, and in the winter to the north. Further to the south, cyclones move in an easterly or south-easterly direction, usually keeping some distance south of the Australian coast. These conditions can conveniently be visualized by considering a typical synoptic chart, shown as Figure 1, supposed to be

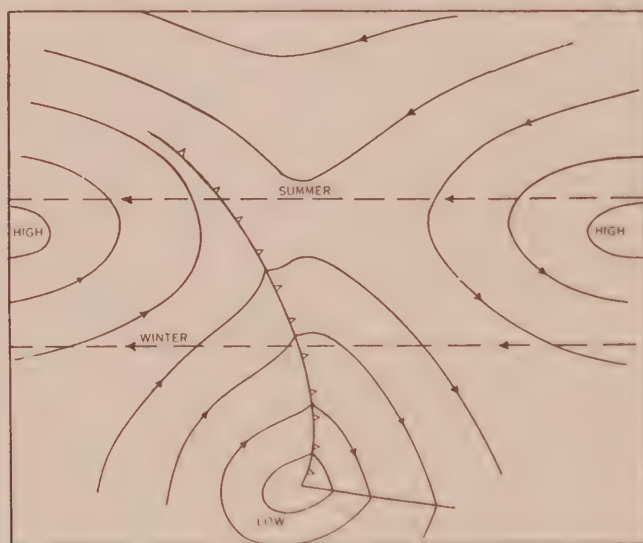


Fig. 1.—Typical synoptic chart in Eastern Australia.

stationary, across which Sydney drifts from right to left. The mean paths of Sydney relative to the chart are shown for summer (January) and winter (July). Thus the synoptic situation at Sydney on a particular occasion can usually be approximately represented by a point on the chart of Figure 1 showing where Sydney was, relative to the chart, at the time concerned.

The most common deviation from this simple situation has been caused, in the period covered by this paper, by tropical lows. These sometimes arose in, or to the west of, central Queensland, and drifted in a south-easterly direction to the Sydney area. Others arrived at the Queensland coast from the north-east and moved down the coast, usually moving out to sea again in the neighbourhood of Sydney. When these lows were present, the anticyclones were often forced much farther south than is usual for the time of the year. These tropical lows have been particularly common in 1950.

(b) Situations Favourable for Rainmaking

On 19 flights, successful experiments have been carried out in which precipitation is believed to have been artificially induced. In 12 of these, the

clouds concerned were deep cumulus and the precipitation reached the ground. In the rest, precipitation released from alto-cumulus evaporated before reaching the ground.

In Figure 2 the synoptic situations for the days on which these experiments were carried out are approximated, as described above, by inserting the position of Sydney on the typical synoptic chart for 0900 hours E.A.S.T. The positions fall in a distinct group around the trough of low pressure between the anti-cyclones. The points are indicated according to season: squares for summer

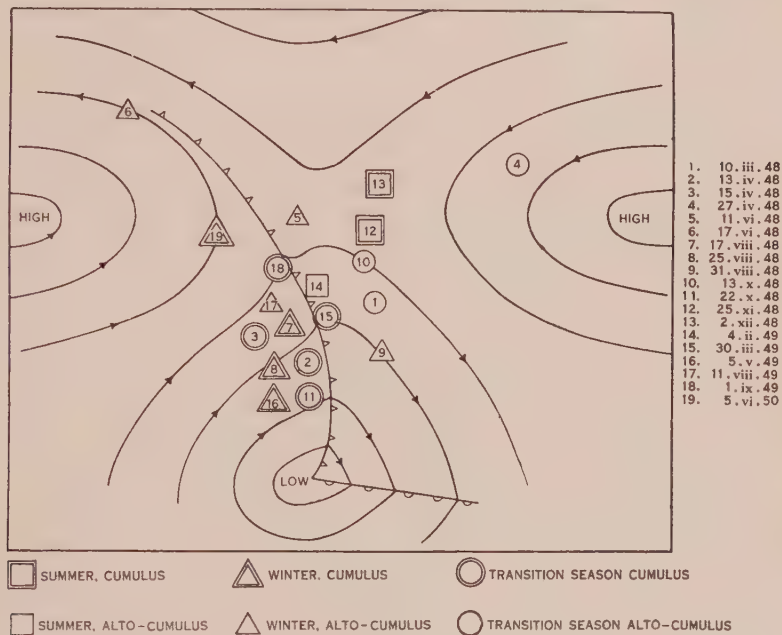


Fig. 2.—Synoptic situations in which successful experiments were made.

(November-February), triangles for winter (May-August), and circles for the transition months (March, April, September, October). Single circles, triangles, or squares represent alto-cumulus, while double symbols represent cumulus from which rain reached the ground. The points fell in different positions in the different seasons, particularly with deep cumulus. These in summer were found in the eastern side of the trough, before the cold front (if any), and about on the latitude of the anticyclones. In winter the successful experiments were carried out on the western side of the trough, usually after the passage of a cold front, and to the south of the latitude of the anticyclones. In the transition months successful experiments were performed both before and after the passage of the trough.

Experiments on alto-cumulus were confined to the winter or transition months. Their grouping was not so pronounced as for cumulus, but they still show a definite tendency to occur near the trough.

In summer the most unstable conditions favouring the growth of deep cumulus usually occur in tropical maritime air with northerly winds after the passage of an anticyclone but before a cold front. In winter, however, unstable

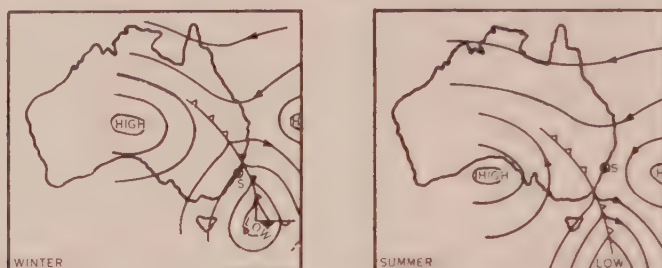


Fig. 3.—Ideal synoptic situations for rainmaking, summer and winter.

conditions with deep cumulus occur most often in cold southern air in south-west winds after the passage of a cold front.

The synoptic situations when natural rain fell in the Sydney area were also approximated in the same way, by inserting the position of Sydney on Figure 1.

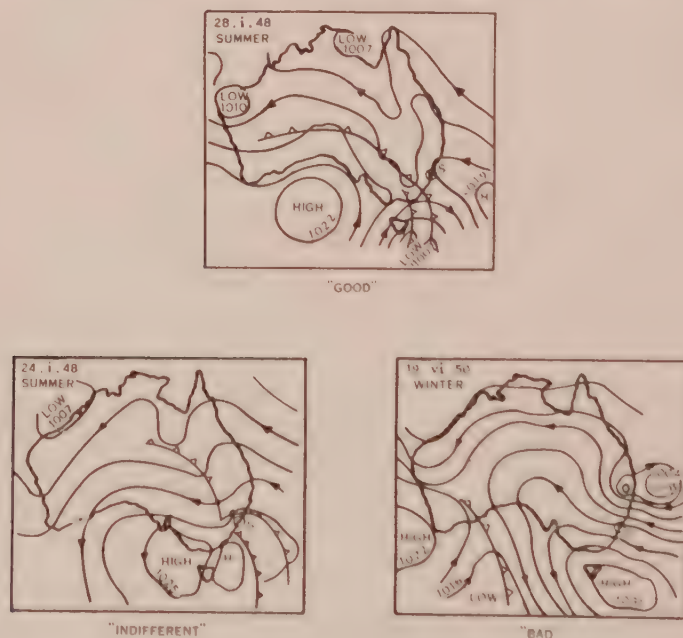


Fig. 4.—Examples of "good", "indifferent", and "bad" situations.

The positions showed a tendency to be most dense in two places: at the trough of low pressure, and to the north of anticyclones. The grouping was, however, not as pronounced as with artificially-induced precipitation. There were also many cases of natural rain associated with tropical lows, particularly in 1950.

(c) Classifications of Situations for Rainmaking

On the basis of Figure 2 the ideal synoptic chart for rainmaking in the Sydney area can be visualized. It varies according to season. In Figure 3, ideal charts are shown for summer and winter. The point *S* represents the position of Sydney. For intermediate months, the conditions to be expected

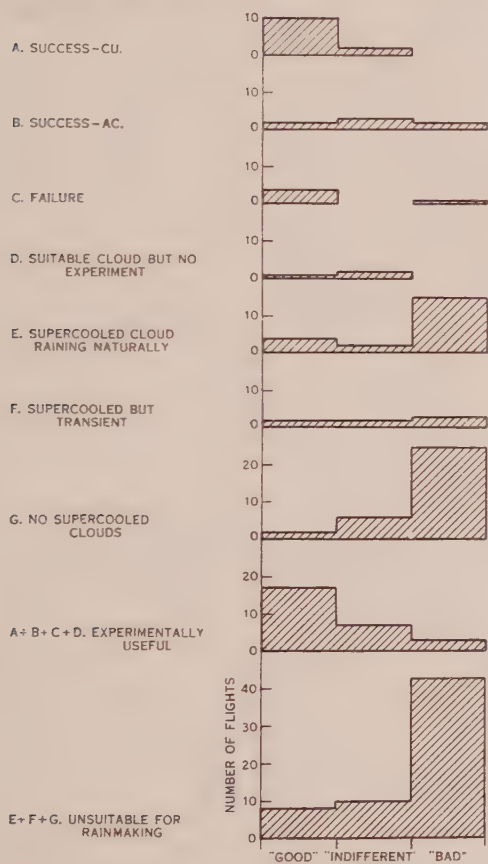


Fig. 5

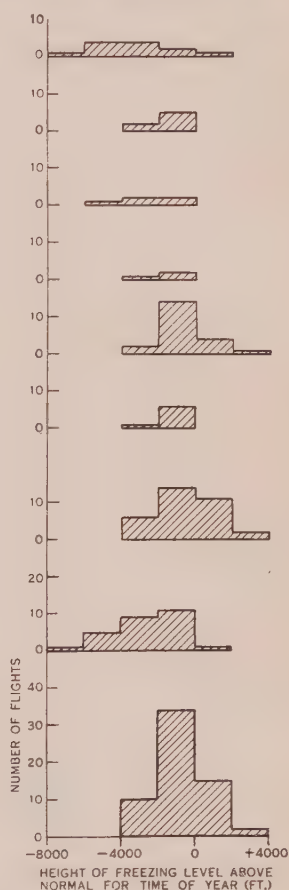


Fig. 6

Fig. 5.—Results in various synoptic situations.

Fig. 6.—Results with various heights of freezing level.

are uncertain, and good results might be achieved with a chart looking like either the summer or winter charts, or anything in between.

Synoptic charts can be classified as to their suitability for rainmaking. This has been done in three categories—good, bad, or indifferent—according as the chart looked like Figure 3 for the appropriate season, entirely unlike it, or intermediate. Samples of the three types of chart are shown in Figure 4.

This process is of course subjective, and the figures which follow might vary slightly if compiled by other investigators. Figure 5 shows the number of occasions when flights in the different categories were made in the different

types of synoptic situation. There is necessarily a correlation between the chance of success and the situation, since a "good" situation is the type in which successes where rain reached the ground were obtained. Figure 5, however, also shows the types of situation in which other results were observed, and

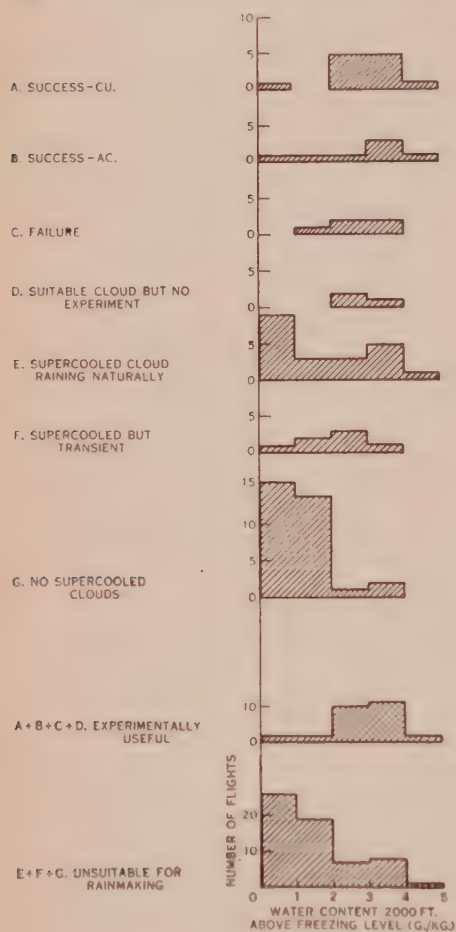


Fig. 7

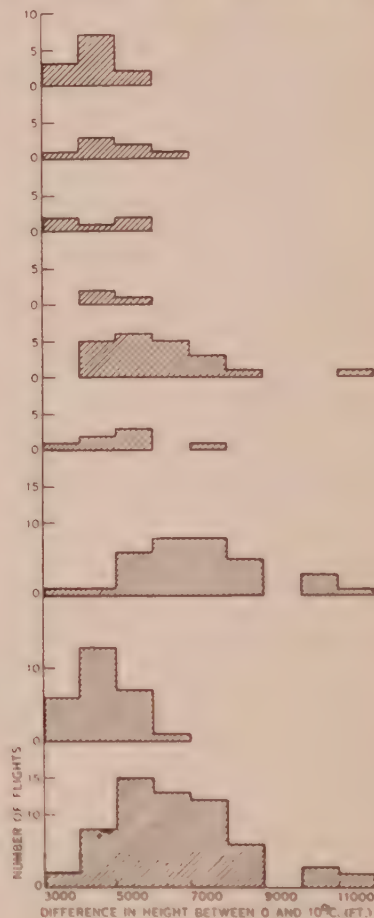


Fig. 8

Fig. 7.—Results with various water contents at 2000 ft. above the freezing level.

Fig. 8.—Results with various lapse rates between 0 and +10 °C.

forms a criterion of the feasibility of sorting situations into those suitable and unsuitable for rainmaking experiments. It appears that although the synoptic chart bears a marked relation to the chance of success, there are other factors which influence the result.

V. HEIGHT OF FREEZING LEVEL

The height of the freezing level varies with the time of year, and also more rapidly within a period of a few days. In investigating the effect of the freezing level on the chance of success, it is desirable to extract the annual variation

and consider the height relative to the normal for the time of the year. Loewe (1945) has given the variation of the freezing level through the year based on aircraft soundings over a period of 10 years at Richmond, New South Wales. This is the aerodrome from which the rainmaking flights were made; it is only 70 miles from Rathmines, where the radiosondes are released. The freezing levels given by Loewe are about 1000 ft. higher throughout the year than those measured by the Rathmines radiosondes in the period covered by this paper.

Sometimes there was an inversion near the freezing level. The "freezing level" quoted in Table 1 and discussed below is then the highest at which the temperature was 0 °C., since experience suggests that this is the level most relevant to rainmaking by dry ice.

Figure 6 gives the number of occasions when flights in the different categories were made when the freezing level occurred at various heights relative to the normal for the time of the year, as given by Loewe. Flights were most often successful with a low freezing level, particularly in the case of deep cumulus clouds where rain reached the ground. Successes on alto-cumulus often occurred when the freezing level was higher.

VI. WATER CONTENT

The average water content in grams per kilogram of air at various heights above the freezing level (as defined in Section V) is presented in Table 2.

TABLE 2
AVERAGE WATER CONTENT AT VARIOUS HEIGHTS FOR FLIGHTS IN EACH CATEGORY (G./KG.)

Height above Freezing Level (thousands of feet)		+6	+4	+2	0	-2	-4	-6	-8
Group									
A	Success—cumulus	1.0	1.5	2.8	3.4	4.1	5.0	6.7	7.6
B	Success—alto-cumulus ..	1.0	1.7	2.6	3.1	4.0	5.0	6.3	8.0
C	Failure	1.3	1.9	2.8	3.4	3.8	4.7	6.1	6.1
D	Suitable clouds but no experiment	1.5	2.0	2.8	3.3	4.0	5.1	5.6	7.1
E	Supercooled but raining ..	0.7	1.0	1.6	2.6	4.0	5.2	6.0	7.1
F	Supercooled but transient	0.8	1.1	2.0	2.6	3.0	5.3	7.4	7.8
G	No supercooled clouds ..	0.7	0.8	0.9	1.3	1.8	3.0	4.7	6.0
A+B+C+D	Experimentally useful ..	1.1	1.7	2.8	3.3	4.0	4.9	6.4	7.4
E+F+G	Unsuitable	0.7	0.9	1.3	1.9	2.7	4.0	5.1	6.6

Where no supercooled clouds were encountered, there was an unusually small water content at all heights. There was also a noticeable deficiency in water content above the freezing level where there were supercooled clouds which were transient or raining naturally.

Krick (1948) has suggested that a high water content at the freezing level is associated with days suitable for seeding clouds in Arizona in summer. Table 2

shows that this is true in the Sydney area, but that the water content 2000 ft. above the freezing level is an even better criterion. Figure 7 shows the number of occasions when the water content at 2000 ft. above the freezing level fell within the limits shown.

It appears strange that on days in category *E*, when all supercooled clouds encountered rained naturally, a smaller average water content should have been measured above the freezing level than on days when supercooled clouds were encountered which did not rain naturally. The large number of occasions in category *E*, when the water content was measured as zero, compared with those in categories *A*, *B*, *C*, *D*, and *F*, suggests that the radiosonde humidity element may break down and register zero if it has been exposed to rain. A different type of radiosonde was introduced during the period covered by the analysis: this did not appear to affect the results. In all these cases, however, the temperature element continued to function, and in all cases some moisture was recorded at lower levels. The fact that zero moisture was frequently recorded when there were no supercooled clouds, but very seldom when there were supercooled clouds which did not rain, justifies some degree of confidence in the readings given in categories other than *E*.

The radiosonde values for water content in flights in category *E* might be disregarded as being unrealistically low. The most water content there could have been is obtained by assuming saturation at the appropriate temperature. Even if this is done, however, the water content is still correlated with the chance of success, due to the large number of days in category *G* when the water content was abnormally low. This assumption also has no effect on the correlation of the observed conditions with the rainmaking index described in Section VIII.

VII. LAPSE RATE BETWEEN 0 AND $+10^{\circ}\text{C}$.

In Figure 8 is shown the number of occasions when flights in the different categories were made with lapse rates in the range shown. Lapse rates are specified by the difference in height, in feet, between the freezing level and the level where the temperature was $+10^{\circ}\text{C}$., taking the highest freezing level and lowest $+10^{\circ}\text{C}$. level when there were more than one. Successes occurred more often with unstable lapse rates, i.e. a small difference in height between 0 and $+10^{\circ}\text{C}$.

VIII. COMBINATION OF PARAMETERS

Four factors have been shown to be related to the chance of success, namely, synoptic situation, height of freezing level, water content 2000 ft. above freezing level, and average lapse rate between the 0 and $+10^{\circ}\text{C}$. levels. Presumably the most suitable days for rainmaking would be those on which all four factors are favourable. A rainmaking index for a day was therefore computed by allotting points according to an arbitrary system, as shown in Table 3.

Figure 9 shows the number of occasions when flights in the different categories occurred on days with given rainmaking indices. Clearly the rainmaking index had a very strong correlation with the chance of success. A

TABLE 3
RAINMAKING INDEX—POINTS SCORE SYSTEM

Synoptic situation	Good Indifferent Bad	Points
		4 2 0
Water content 2000 ft. above freezing level	Over 2 g./kg. ..	3
	Under 2 g./kg. ..	0
Height of freezing level relative to normal (ft.)	Over 2000 below ..	3
	Under 2000 below	1
	Above	0
Lapse rate (ft.) ; between 0 and +10 °C.	Under 4500 ..	4
	4500 to 5500 ..	2
	Over 5500 ..	0

comparison of the correlation between the chance of success, the four meteorological factors and the rainmaking index is given in Table 4.

TABLE 4
COMPARISON OF CORRELATION OF CHANCE OF SUCCESS WITH RAINMAKING INDEX
AND INDIVIDUAL PARAMETERS

Conditions in which Flights were made	Flights per Experimentally Useful Flight	Flights per Success when Rain Reached the Ground
All flights	$\frac{88}{27} = 3.3$	$\frac{88}{12} = 7.3$
When situation was "good" ..	$\frac{25}{17} = 1.5$	$\frac{25}{10} = 2.5$
When water content > 2g./kg. 2000 ft. above freezing level	$\frac{39}{23} = 1.7$	$\frac{39}{11} = 3.5$
When freezing level > 2000 ft. below normal	$\frac{26}{15} = 1.7$	$\frac{26}{9} = 2.9$
When height difference between 0 and +10° < 5000 ft.	$\frac{29}{19} = 1.5$	$\frac{29}{10} = 2.9$
When rainmaking index exceeded 8 ..	$\frac{25}{23} = 1.1$	$\frac{25}{11} = 2.3$
When rainmaking index exceeded 10 ..	$\frac{13}{13} = 1.0$	$\frac{13}{10} = 1.3$

It is not likely that all the four factors which have been considered are entirely independent of each other; nor is it possible, in view of the limited

accuracy of the data, to establish with any precision the relations, if any, between them. Each was plotted in turn against the others, in each case revealing a small degree of correlation and a large scatter. An attempt was made to find out whether any of the four factors did not contribute to the value of the rain-making index, by allotting points for each combination of three factors. The

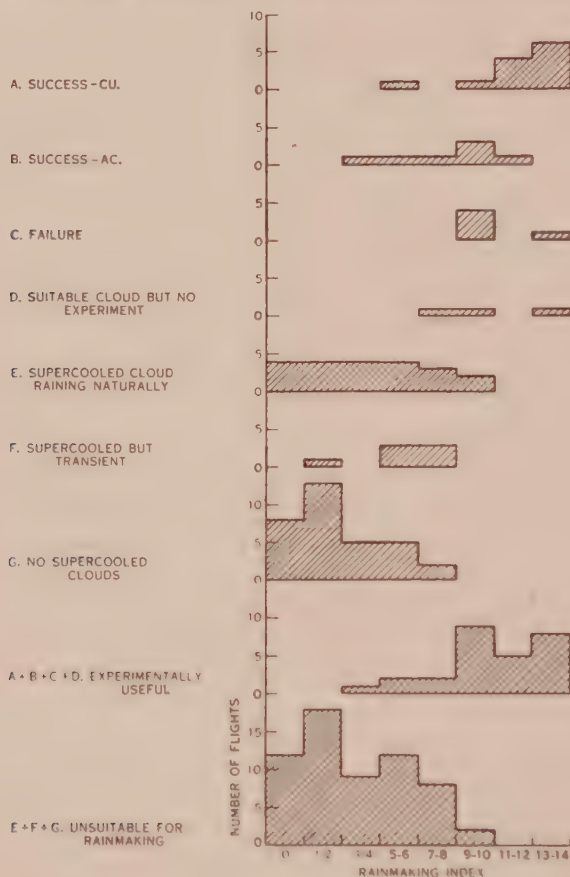


Fig. 9.—Results with various "rainmaking indices".

result in each case was an index which had less correlation with the chance of success than the index involving all four. It is concluded that each of the four factors contributed, to some extent, to the chance of success, independently of the others.

IX. ESTIMATE OF NUMBER OF OPPORTUNITIES FOR RAINMAKING

Since on the days when flights were made the rainmaking index formed a very good measure of the chance of success, it seemed reasonable to apply it to all days during the period of the analysis in order to estimate the total number of opportunities for rainmaking. This was done in the case of experiments with deep cumulus, when rain should have reached the ground, by considering the

ratio of flights to successes on days with over 10 points (1.3) and with 5 to 10 points (18), taken from Figure 9. The results are shown in Table 5: an "opportunity" means a day on which it should have been possible to cause rain to reach the ground. There was no significant difference in the number of

TABLE 5
ESTIMATE OF NUMBER OF OPPORTUNITIES FOR RAINMAKING

Rainmaking index	1948				1949				1950 (January to June)			
	0-4	5-10	11-14	Total	0-4	5-10	11-14	Total	0-4	5-10	11-14	Total
Number of cases for which data are available ..	142	134	40	316	163	140	28	331	106	51	5	162
Estimate of total number of cases ..	165	155	46	366	180	154	31	365	118	57	6	181
Estimated number of opportunities for rainmaking ..	0	9	36	45	0	9	24	33	0	3	5	8

opportunities in the different seasons, but there was a marked variation from year to year: in January to June 1950 there were only eight opportunities, compared with 22 and 15 in a similar period in 1948 and 1949 respectively. The average natural rainfall over an area within 60 miles of Sydney in 1948 was 31.9 in. (0.93 times the normal): in 1949 it was 48.6 in. (1.41 times normal) and in January to June 1950 it was 46.9 in. (2.56 times normal). Thus there were most opportunities for rainmaking when the natural rainfall was least. It is likely that in other areas the reverse would apply, if, for example, there were no rain owing to absence of clouds. Clearly it is desirable that experiments in the Sydney area be continued in at least one season when the rainfall is substantially less than normal.

No information is available as to the number of occasions when it would have been possible to cause rain to fall on any given place. Sometimes there were large numbers of similar clouds, and one was selected at random for experiment. In these cases, given suitable navigational facilities, it should have been possible to cause rain to fall in the desired place. In other cases experiments were made on the only suitable cloud for many miles. The number of opportunities per year for making rain in any one place was clearly considerably less than that quoted above.

X. FORECASTING OPPORTUNITIES FOR RAINMAKING

The analysis reported above has correlated the chance of success in rainmaking on a given day with the synoptic chart for 0900 hours and the radiosonde

sounding at Rathmines at 1800 hours of the same day. In New South Wales the synoptic charts are drawn every 3 hours, but during the period under discussion there was only one radiosonde daily. Consequently a forecast of the chance of success (and a decision of whether to fly) would have to be based on the synoptic chart for 3 hours before the intended time of take-off, and the radiosonde sounding for the day before. An attempt was made to determine the relevance of these data by computing the rainmaking index as described in Section VIII, but using the 0900 hours chart and the previous day's sounding. The result had slightly less correlation with the chance of success than had the synoptic situation alone. Consequently with the data which existed at the time, forecasts could best be based on the synoptic chart alone, and the degree of success of such forecasting can be judged from Figure 5: about $2\frac{1}{2}$ flights per success, when rain reached the ground, could be expected if flights took place only when the situation was "good".

It is likely that a much improved forecast could be made if a radiosonde sounding were made immediately before a proposed flight. If the resulting forecast had as good a chance of correlation with the chance of success as the rainmaking index, the success of the forecast could be judged from Figure 9, and the number of flights per success would be 1.3 if flights took place when the index exceeded 10. In the period January 1948 to June 1950 there would have been about 65 successes, and flights would have taken place on 160 or 85 days respectively, based on these two methods of forecasting.

XI. CONCLUSIONS

It has been shown that the incidence of clouds suitable for artificially stimulating precipitation depends on certain specific meteorological conditions. The number of occasions per year when suitable conditions occurred has been estimated. The above analysis applies only to the Sydney area, and the conclusions cannot necessarily be expected to hold good elsewhere.

The number of flights and the accuracy of the data are too low to allow precise deductions to be drawn. The analysis should be repeated when further results are available.

It is clear, however, that near Sydney the number of opportunities for rainmaking is small. Consequently if practical rainmaking were to be attempted, three conditions would need to be fulfilled. The rainmaking aircraft must be prepared to fly whenever the weather is suitable: flights planned without regard to the weather yield very poor returns. Every effort must be made to provide adequate specialized forecasts: this can probably be done if additional radiosonde soundings are made. Finally, when the opportunity does arise, several flights per day should be made in order to obtain the utmost benefit from it.

XII. ACKNOWLEDGMENTS

The author expresses his gratitude to the Weather Bureau, Sydney, for the meteorological information, and to the Royal Australian Air Force, in whose aircraft the flights were made.

XIII. REFERENCES

- LOEWE, F. (1945).—Bull. Bur. Met. Aust. No. 33, p. 13.
- KIDSON, E. (1925).—Bull. Bur. Met. Aust. No. 17, pp. 11-13.
- KRAUS, E. B., and SQUIRES, P. (1947).—*Nature* **159**: 489-92.
- KRICK, I. P. (1948).—Report on meteorological aspects of artificial nucleation in the Phoenix (California) area in summer. (Private publication.)
- SMITH, E. J. (1949).—*Aust. J. Sci. Res. A* **2**: 78-91.
- SMITH, E. J. (1950).—*Aust. J. Sci. Res. A* **3**: 214-23.
- SQUIRES, P., and SMITH, E. J. (1949).—*Aust. J. Sci. Res. A* **2**: 232-45.

THE STOPPING POWER OF ILFORD C₂ EMULSION

By A. J. DYER*

[Manuscript received October 24, 1951]

Summary

The stopping power of Ilford C₂ nuclear emulsion under fixed water content conditions has been determined for protons in the energy range 1.73–5.26 MeV. with an accuracy of 1 per cent. The variation of stopping power with water content has been calculated and an appropriate correction applied. It has been found that the values obtained in this experiment are in substantial agreement with those of Lattes, Fowler, and Cier, and of Rotblat. A consistency in stopping power of better than 0.3 per cent. has been observed between several batches. Some data on the straggling of protons in nuclear emulsion are presented.

I. INTRODUCTION

A very complete analysis of the existing range-energy data for all Ilford nuclear emulsions has been made by Wilkins (1951), who has shown that the absorption of water by nuclear emulsion takes place according to a simple law of mixtures. He has revised values of the differential stopping power of the emulsion components and has deduced the particular water content conditions in Rotblat's (1950) calibration experiment. Wilkins used this data to develop a range-energy table for α -particles. The proton range-energy relation for energies greater than 0.75 MeV. was then derived from the expression $R_p(E) = 1.0069R_\alpha(3.973E) - 1.38\mu$. The corrections necessary for various emulsion loadings have been calculated, and approximate data for the water content correction have been given. The tabulated results are found to be in good agreement with the work of Lattes, Fowler, and Cier (1947), Bradner *et al.* (1950), Nereson and Reines (1950), and Panofsky and Fillmore (1950).

Rotblat (1951) has corrected his earlier measurements and extended them to higher energies. These later results are in good agreement with those of Catala and Gibson (1951) who used an almost identical method. Wilkins (1951) finds that the new calibration can be fitted adequately with his semi theoretical calculation with a slight adjustment of the water content parameter.

The purpose of this paper is to present the results of stopping-power determinations for protons in C₂ emulsion in the energy interval 1.7–5.3 MeV. The effect of water content on emulsion stopping power has been calculated and an appropriate correction applied to these results.

II. METHOD

(a) Exposure of Plates

High energy protons were obtained by deuteron bombardment of thin Li⁶ and C¹² targets employing the nuclear reactions Li⁶(d,p)Li⁷ and C¹²(d,p)C¹³.

* Physics Department, University of Melbourne.

The 500 keV. deuteron beam used was obtained from a 1 MeV. electrostatic generator equipped with voltage stabilization and current regulation of the magnetic analyser, which maintained the deuteron energy constant to within 2 keV. of its nominal value.

The 50 μ C₂ plates were exposed in a camera, described previously (Martin *et al.* 1949), which allows the emergent particle to be recorded in the angular range 13–167° relative to the incident beam. Thin targets of Li⁶ and C¹², approximately 5 keV. thick, were deposited on aluminium leaf.

Between the target and the plate a thickness of red cellophane was interposed to absorb scattered deuterons. The protons in the forward direction also had to traverse the thin aluminium leaf forming the target backing. For the experiment 15 plates from a variety of batch numbers were used, the average exposure for each plate being approximately 45 microampere minutes.

(b) *Measurement of Tracks*

In the camera the plate lies in a plane parallel to the deuteron beam but offset 1 cm. from it. The tracks were measured in an area extending on each side of the various angles chosen and between two circles of 4.0 and 4.5 cm. radius, the common centre of which was directly above the target spot. Counts showed that the inverse square law was valid in this region. The required angles and circles were marked on the plate by scribing the emulsion with a fine needle point, after which the plates were cut into quarters to facilitate handling on the microscope. The angles could be determined to within 0.25° and the maximum angle subtended by the scanned area of the plate at the target was 1°. The average angle of incidence (β) of the protons to the plane of the plate was 13.3°.

A binocular measuring microscope, which has been described previously (Dunbar and Bower 1949), was employed for measuring the projected ranges (L) of the tracks. By taking the true range as $L \sec \beta$ we avoided making a correction for shrinkage of the emulsion in the z -direction. A $\times 42$ oil objective and $\times 10$ eyepieces were used for the lithium protons, and a $\times 92$ oil objective and $\times 15$ eyepieces for the carbon protons. The latter magnifications were chosen to make the required tracks extend over some two-thirds of the field of view. This was considered a more satisfactory practice than that of using the highest magnification in all cases and stepping off along the track. The Li⁶ and C¹² protons were measured using two different eyepiece scales which were each internally consistent to 0.1 per cent. A fixed interocular distance and draw-tube position were maintained to prevent any change in magnification; however, every time measurements were undertaken the eyepiece scale used was checked against a Grayson ruling which had been calibrated by a Hilger travelling microscope. This procedure allowed the projected range to be measured with an accuracy of 0.3 per cent.

The tracks selected for measurement were straight tracks and tracks only requiring a small correction for multiple scattering. In many cases multiple scattering causes a marked deviation near the end of the track where the particle has a small velocity. Tracks of this type which required not more than half a

division correction for this deviation in a total range of the order of 60 divisions were measured. Frequent reorientation of the eyepiece scale does not permit accurate measurement of tracks showing gradual curvature for the major portion of their length and these were rejected, together with tracks which showed abrupt changes in direction. Although the above criteria provided adequate selection in the x - y plane, no simple scheme could be adopted for selection in the z -direction except that tracks which showed severe changes in dip were rejected. The number of tracks excluded by this method of selection amounted to about 15 per cent. of the total yield.

On the average, 670 tracks were measured for each proton group giving a total of 10,000 tracks for the whole experiment. It was found that two determinations of several mean ranges by separate observers were consistent to 0.2 microns.

(c) Water Content Correction

Because of the slow pumping speed within the camera, it was not convenient to vacuum-dry the plates completely. The procedure of maintaining the plates in the vacuum for 20 minutes and then bombarding the target for 40 minutes was therefore adopted in all cases. The necessary correction for water content was then made in the following manner.

In Section IV, the variation of emulsion stopping power with density ρ is calculated, using Wilkins's (1951) value of 3.958 g. cm.³ for a completely vacuum-dried emulsion. Wilkins demonstrates that the emulsion-water mixture consists of a simple addition of volumes. This is the basis of equation (1) of Section IV, which for the present purpose is more conveniently expressed in the form

$$\rho = \frac{4.18(1+m)}{1+4.18m},$$

where m g. is the mass of water absorbed by 1 g. of absolutely dry emulsion, which has a density of 4.18 g. cm.⁻³.

Let m_0 , m_1 , and m_2 respectively refer to the values of m for a completely vacuum-dried plate, before insertion in the camera, and during the period of a run. A value of 0.0179 g. can be calculated from Wilkins's data for m_0 .

Three plates were exposed in the camera under the same conditions as the calibration plates, and by weighing a mean value of 5.5×10^{-3} g. was found for $m_1 - m_2$. Three additional plates were exposed in a fast pumping system and by weighing to constant weight we find $m_1 = 0.0337$ g. and hence $m_2 = 0.0282$ g.

Substituting this value for m_2 in the above expression gives $\rho_2 = 3.84 \pm 0.02$ g. cm.³, the limits of the error being given by the standard deviation of three measurements and allowing for the gradual removal of water during the period of bombardment. The consistency of the results confirms the opinion that the moisture content of the emulsion before insertion in the camera, i.e. when wrapped and stored, is reasonably constant.

A correction curve for a density of 3.84 g. cm.⁻³ was then set up from Figure 3 in the form: percentage correction plotted against range in microns. The

existing range-energy relation could be used with confidence because of the extremely slow energy-dependence of the correction. All observed mean ranges were now corrected according to this curve.

(d) *Window Correction*

In calculating the emulsion stopping power a correction was necessary since, before reaching the plate, the protons traversed a thickness of cellophane. It was also necessary to test the latter for uniformity and a series of checks showed that any variation in cellophane thickness was less than ± 2 per cent.

TABLE 1

Energy E_p (MeV.)	Range in Emulsion (μ microns)	Range in Air (cm.)	Stopping Power
5.261 ± 0.007	187.6 ± 1.0	37.0 ± 0.05	1972 ± 12
5.013 ± 0.009	173.6 ± 0.9	34.1 ± 0.05	1964 ± 12
5.014 ± 0.010	173.1 ± 0.9	34.1 ± 0.07	1970 ± 16
4.756 ± 0.011	159.1 ± 0.8	31.1 ± 0.09	1955 ± 16
4.757 ± 0.012	158.8 ± 0.8	31.1 ± 0.10	1958 ± 16
4.607 ± 0.011	150.5 ± 0.8	29.4 ± 0.09	1953 ± 16
4.489 ± 0.013	145.0 ± 0.8	28.1 ± 0.10	1938 ± 16
4.490 ± 0.014	143.9 ± 0.8	28.1 ± 0.11	1953 ± 16
4.336 ± 0.013	136.2 ± 0.7	26.5 ± 0.11	1946 ± 18
4.212 ± 0.016	129.5 ± 0.7	25.15 ± 0.14	1942 ± 20
4.200 ± 0.007	129.2 ± 0.7	25.00 ± 0.05	1934 ± 14
4.049 ± 0.015	121.8 ± 0.7	23.55 ± 0.13	1933 ± 20
3.907 ± 0.009	114.6 ± 0.6	22.20 ± 0.07	1937 ± 16
3.597 ± 0.011	100.5 ± 0.5	19.25 ± 0.08	1915 ± 18
3.267 ± 0.013	86.0 ± 0.5	16.30 ± 0.09	1895 ± 20
2.913 ± 0.008	71.2 ± 0.4	13.35 ± 0.04	1875 ± 16
2.683 ± 0.006	62.3 ± 0.4	11.60 ± 0.04	1862 ± 16
2.517 ± 0.010	56.4 ± 0.3	10.40 ± 0.05	1844 ± 20
2.254 ± 0.008	47.4 ± 0.3	8.65 ± 0.05	1825 ± 22
2.062 ± 0.012	40.8 ± 0.2	7.38 ± 0.06	1809 ± 28
1.748 ± 0.010	31.8 ± 0.2	5.62 ± 0.06	1767 ± 30

The correction that had to be applied to the emulsion stopping power to allow for the one thickness of cellophane was obtained by exposing further plates for both the lithium and carbon reactions with two and three thicknesses of cellophane in position. Measurement of the latter plates allowed the residual mean range of protons at the angles concerned to be determined after traversing two and three thicknesses of cellophane. The change in range ($\Delta\mu$ microns) produced by each additional window was plotted against the residual range (μ microns) as shown in Figure 1. The line shown (Fig. 1) is a theoretical ($\Delta\mu$, μ) curve calculated from the known composition and surface density of the cellophane (C₆H₁₀O₅) and differential stopping-power data of Wilkins (1951). The theoretical

curve is seen to be adequately verified by the experimental results, and the necessary corrections to obtain the mean range without window were applied from this curve.

From the above theoretical calculation and the accepted range-energy relations for protons in air (Livingston and Bethe 1937; Bethe 1950) it was possible to determine the energy lost by a proton traversing the cellophane window. The energies corresponding to the various proton groups measured

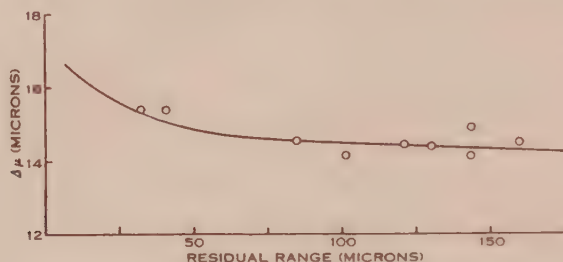


Fig. 1.—Loss of range $\Delta\mu$ microns caused by cellophane window, compared with calculated curve.

for different window thicknesses could then be determined. We have estimated that this calculation has an error of about 0.002 MeV. The results so obtained are included in Table 1.

III. RESULTS

The mean range was determined at various angles α in the plane of the plate relative to a reference line parallel to the incident beam. Values of α of 0, 55, 110, and 180° were chosen for the Li^6 protons and 90 and 180° for the C^{12} protons, the actual angle of emission of the particles relative to the deuteron beam being $\arccos(\cos \alpha \cos \beta)$.

In this experiment, mean range refers to the distance reached by half the number of particles following the definition of Livingston and Bethe (1937).

Small corrections of 0.6 μ and 0.4 μ were necessary for the proton ranges observed at 0 and 55° respectively, due to absorption in the aluminium target leaf.

The proton energies were derived from the relation

$$M_n Q = (M_n + M_p) E_p - (M_n - M_d) E_d - 2 \cos \theta (M_d M_p E_d E_p)^{1/2}.$$

Here M_n , M_p , and M_d are the masses of the residual nucleus, proton, and deuteron respectively. E_p and E_d in MeV, are the corresponding particle energies and θ is the angle of emission relative to the incident beam in laboratory coordinates. Q represents the energy release which for the Li^6 reaction is 5.019 ± 0.007 MeV, and for the C^{12} reaction 2.716 ± 0.005 MeV. (Strait *et al.* 1951).

The results of the present experiment are shown in Table 1, which records the emulsion stopping power, range in air, and range in emulsion for different proton energies. The energy uncertainty in E_p (Table 1) has been calculated allowing for errors in E_d , θ , and Q whilst the uncertainty in μ is determined by

the precision of the eyepiece-scale calibration, the accuracy of the cellophane window correction, the water content correction, and the statistical variation of the range distribution. Both the errors in E_p and μ influence the accuracy of stopping-power values.

In Figure 2, we have plotted the stopping-power curve obtained by Rotblat together with the results of this experiment and those of other workers. The theoretical calculation of Wilkins is almost identical with that of Rotblat and

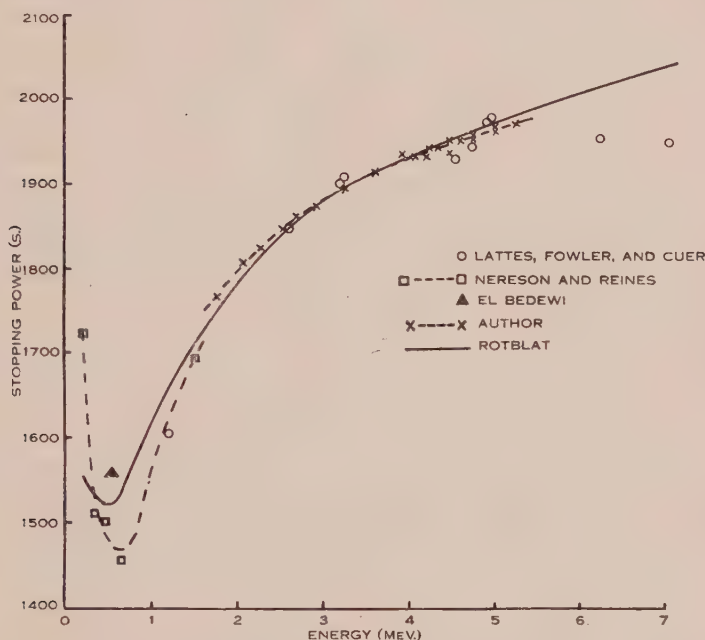


Fig. 2.—Variation of stopping power of Ilford C₂ emulsion with energy.

has been omitted to avoid confusion. The dotted line is a curve of best fit to our points, and it is seen that no point is more than 0.5 per cent. from this mean curve. In the range 3 to 4 MeV. the two curves are identical but below 3 MeV. our stopping-power values are 0.5 per cent. higher than those of Rotblat, and above 4 MeV. our points are 0.5 per cent. lower. This is hardly significant in view of the accuracy of both experiments. Both curves are in satisfactory agreement with Lattes, Fowler, and Cier (1947) but it is noted that there is a considerable difference in the results of the various workers below 1.5 MeV. Although the points of Nereson and Reines (1950) and of El Bedewi (1951) were obtained with plates not exposed *in vacuo*, the discrepancy cannot be resolved on the basis of a water content correction. It is most likely caused by the difficulty of measuring precisely ranges of only a few microns.

Plates from four batches were used in this experiment and we have found a consistency in stopping power of better than 0.3 per cent. from batch to batch.

IV. WATER CONTENT CALCULATION

Wilkins (1951) has shown that the absorption of water by nuclear emulsion takes place according to the law of simple mixtures and expresses the condition of the emulsion in terms of the volume of water ω cc. absorbed by 1 cc. of absolutely dry emulsion (density 4.18 g. cm.^{-3}). Thus the absorption of ω cc. of water by 1 cc. of absolutely dry emulsion results in $(1 + \omega)$ cc. of emulsion of density ρ , where ρ is given by

$$\rho = (4.18 + \omega)/(1 + \omega) \text{ g. cm.}^{-3}. \quad (1)$$

To calculate the effect of water content use is made of the formula for the differential stopping power of a mixture

$$dS = \frac{dR}{d\mu} = \frac{\sum_i s_i V_i}{\sum_i V_i}, \quad (2)$$

where R, μ refer to range in air and emulsion respectively. V_i is the volume of the component i of the mixture and s_i is the differential stopping power of component i . We are at liberty to make $\sum_i V_i$ equal to 1 cc. and thus for an absolutely dry emulsion

$$dS_0 = \sum_i s_i V_i, \quad (3)$$

and for an emulsion with water content ω

$$dS_\omega = \frac{dS_0 + \omega dS_w}{1 + \omega}, \quad (4)$$

where dS_w is the differential stopping power of water.

We now propose to calculate the correction Δ_ω necessary to convert a stopping power S_ω to a value S_{ω_0} corresponding to a standard water content ω_0 .

Writing $k_\omega = \frac{1}{dS_\omega}$

$$\mu_\omega = \int_0^R \frac{1}{dS_\omega} dR = \int_0^R k_\omega dR. \quad (5)$$

And from (4)

$$k_\omega = \frac{1 + \omega}{1 + \omega_0} \cdot \frac{dS_0 + \omega_0 dS_w}{dS_0 + \omega dS_w} \cdot k_{\omega_0},$$

or neglecting second and higher order terms in ω

$$k_\omega = k_{\omega_0} \frac{1 + \omega}{1 + \omega_0} \left\{ 1 + (\omega_0 - \omega) \frac{dS_w}{dS_0} \right\}. \quad (6)$$

If this expression is integrated according to (5),

$$\Delta_\omega = \frac{S_{\omega_0} - S_\omega}{S_\omega} = \frac{\mu_\omega - \mu_{\omega_0}}{\mu_{\omega_0}} = \frac{\omega - \omega_0}{1 + \omega_0} \{ 1 - (1 + \omega) I / \mu_{\omega_0} \}, \quad (7)$$

where

$$I = \int_0^R \frac{dS_w}{k_{\omega_0} dS_0} dR.$$

It can be seen from (7) that the factor $(\omega - \omega_0)/(1 + \omega_0)$ describes the swelling of the emulsion causing a decrease in the stopping power, and the term $(1 + \omega)I/\mu_{\omega_0}$

which is approximately equal to 0.5 allows for the increase in stopping power due to the added material.

The latter term was obtained by a numerical integration using the differential stopping-power data of Wilkins and taking $\omega_0 = 0.075$ which is the value he found to fit the revised results of Rotblat. It is estimated that even for the extreme

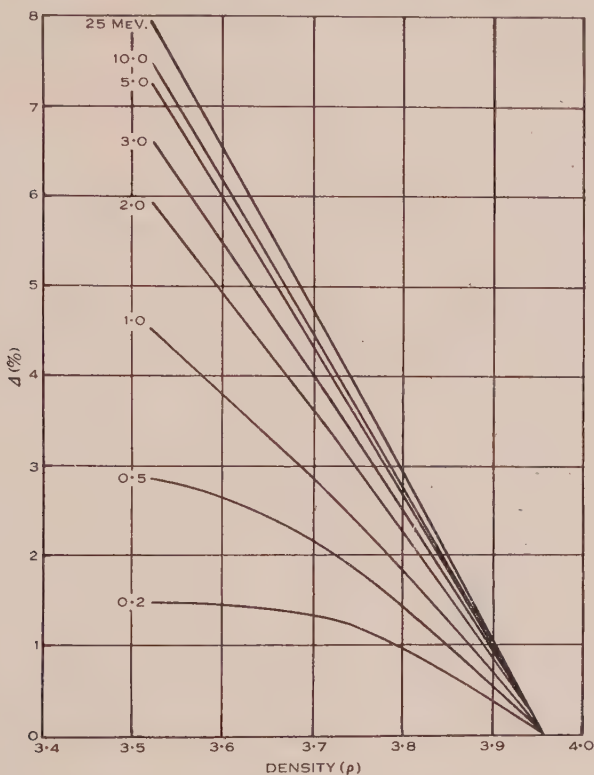


Fig. 3.—Effect of water content on stopping power, expressed as percentage correction necessary to convert to a vacuum-dried emulsion.

case of $\omega = 0.240$ cc. ($\rho = 3.56$ g. cm.⁻³) the error incurred in Δ by taking a first order approximation in ω is less than 1 per cent.

The correction Δ in per cent. is plotted against emulsion density in Figure 3 for convenient energy values. At a relative humidity of 60 per cent., ρ is approximately 3.78 g. cm.⁻³, and it can be seen that the stopping power of the emulsion is 2–3 per cent. lower than for a standard vacuum-dried emulsion.

V. PROTON STRAGGLING

The straggling coefficient s has been calculated from $s = \Delta\mu/\mu$ where $\Delta\mu$ is half the width at half the maximum height of the range distribution curve, for those particle groups which traversed only one cellophane window. The values of s so obtained were corrected for “angular” straggling and are plotted in

Figure 4, together with the results of Nereson and Reines (1950), El Bedewi (1951), and Rotblat (1950).

Our values agree with Rotblat above 2.3 MeV., but support the lower values found by Nereson and Reines below this region, and are not inconsistent with the point obtained by El Bedewi. It must be concluded that the additional

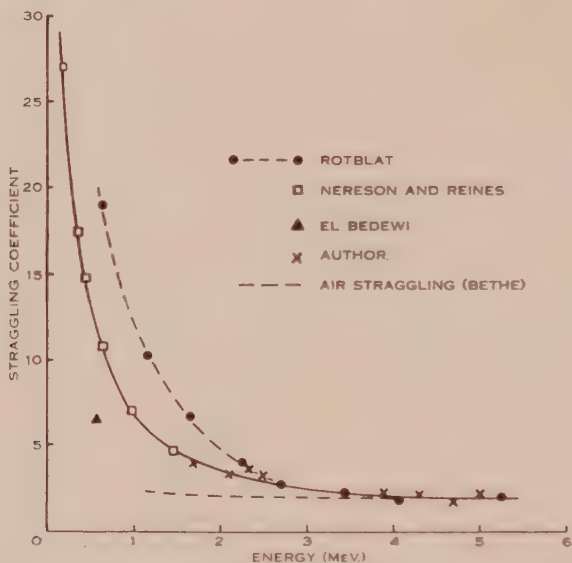


Fig. 4.—Straggling coefficient for protons plotted against energy.

straggling factors for which Rotblat corrected were somewhat greater than he estimated. It is seen that the straggling of protons in C_2 nuclear emulsion is greater than that in air as given by Livingston and Bethe (1937) below an energy of 4 MeV.

VI. ACKNOWLEDGMENTS

Thanks are due to Professor L. H. Martin for his continued interest during this investigation. The work was carried out under the direction of Dr. J. C. Bower and the author has appreciated his helpful advice. A considerable amount of the microscope work was carried out by Mr. W. Dowell, and Mr. F. Hirst assisted in making the exposures. The author is grateful to Mr. J. R. Bird for depositing the Li^6 targets, and to the staff of the Defence Research Laboratories for ruling the eyepiece scales and calibrating the Grayson ruling.

VII. REFERENCES

- BETHE, H. A. (1950).—*Rev. Mod. Phys.* **22** : 213.
 BRADNER, H., SMITH, F. M., BARKAS, W. H., and BISHOP, A. S. (1950).—*Phys. Rev.* **77** : 462.
 CATALA, J., and GIBSON, W. M. (1951).—*Nature* **167** : 551.
 DUNBAR, D. N. F., and BOWER, J. C. (1949).—*J. Sci. Instrum.* **26** : 132.
 EL BEDEWI, F. A. (1951).—*Proc. Phys. Soc. A* **64** : 584.
 LATTES, C. M. G., FOWLER, P. H., and CUER, P. (1947).—*Proc. Phys. Soc. A* **59** : 883.

- LIVINGSTON, M. S., and BETHE, H. A. (1937).—*Rev. Mod. Phys.* **9** : 245.
- MARTIN, L. H., BOWER, J. C., DUNBAR, D. N. F., and HIRST, F. (1949).—*Aust. J. Sci. Res. A* **2** : 25.
- NERESON, N., and REINES, F. (1950).—*Rev. Sci. Instrum.* **21** : 534.
- PANOFSKY, W. K. H., and FILLMORE, F. L. (1950).—*Phys. Rev.* **79** : 57.
- ROTLAT, J. (1950).—*Nature* **165** : 387.
- ROTLAT, J. (1951).—*Nature* **167** : 550.
- STRAIT, E. N., VON PATTER, D. M., BUECHNER, W. N., and SPERDUTO, A. (1951).—*Phys. Rev.* **81** : 747.
- WILKINS, J. J. (1951).—A.E.R.E. Rep. G/R664.

INFLUENCE OF THE VOLUME FRACTIONS OF THE PHASES ON THE DEFORMATION OF ($\alpha + \beta$) BRASS

By L. M. CLAREBROUGH* and G. R. PERGER*

[Manuscript received September 18, 1951]

Summary

The relative deformations of the phases were studied, after 20, 40, 60, and 80 per cent. reductions in area by wire drawing in duplex brasses of various compositions, using as a measure of the degree of deformation the temperature at which recrystallization began. In alloys containing more than about one-third by volume of the hard β phase, crystals of both phases deformed to the same extent, but in alloys containing 10 or 20 per cent. by volume of the β phase, the heavier deformation always occurred in the soft α phase. The mode of recrystallization of the phases varied with the degree of deformation and the volume proportion of the phases.

I. INTRODUCTION

The first determination of the relative deformation of the phases in duplex alloys was made by Unckel (1937) by studying the changes in shape of the particles (each of which consisted of a number of crystals). However, the method cannot be applied generally, since, in most duplex alloys, when the particles of the phases are observed microscopically on polished sections, their shapes are very irregular and some parts of a second phase particle may be deformed quite heavily without very much change in the shape of the particle as a whole. Therefore, the method adopted here is to use the temperature at which a phase just starts to recrystallize (called recrystallization temperature in this paper) as a measure of the extent of deformation of that phase.

The object of the work is to show how the deformations of the crystals of the α and β phases in duplex brass vary with the volume fractions of the phases and with the total deformation of the alloy. Previous work on silver-magnesium alloys (Clarebrough 1950) was confined to one deformation only.

II. EXPERIMENTAL

The brasses used were prepared from copper and zinc each having a purity of 99.99 per cent. The alloys were chill cast, machined to 0.45 in. diameter, and then swaged to 0.21 in. diameter with one intermediate annealing treatment. For the final heat treatment the alloys were annealed at 600 °C. for 24 hours, cooled to 300 °C. at a rate of 30 °C. per hour, and held at this temperature for 4 days before cooling to room temperature at a rate of 30 °C. per hour. This final treatment was necessary in order to avoid complications arising from non-equilibrium structures (Clarebrough 1950) and from the superstructure in the β phase (Honeycombe and Boas 1948). For all these heat treatments the alloys were sealed in evacuated Pyrex glass tubes.

* Division of Tribophysics, C.S.I.R.O., University of Melbourne.

Six alloys were used, the volume fractions of the β phase being 0.1, 0.2, 0.35, 0.5, 0.6, and 0.7. These alloys were deformed, by drawing through

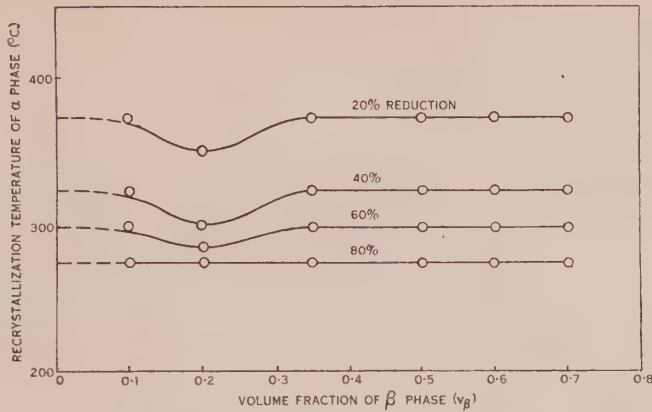


Fig. 1.—Influence of the volume fractions of the phases on recrystallization temperature of α phase after various reductions in area by wire drawing.

tungsten carbide dies, to 20, 40, 60, and 80 per cent. reduction in area. Half-inch lengths of the drawn wires were then annealed in a salt bath for 30 minutes

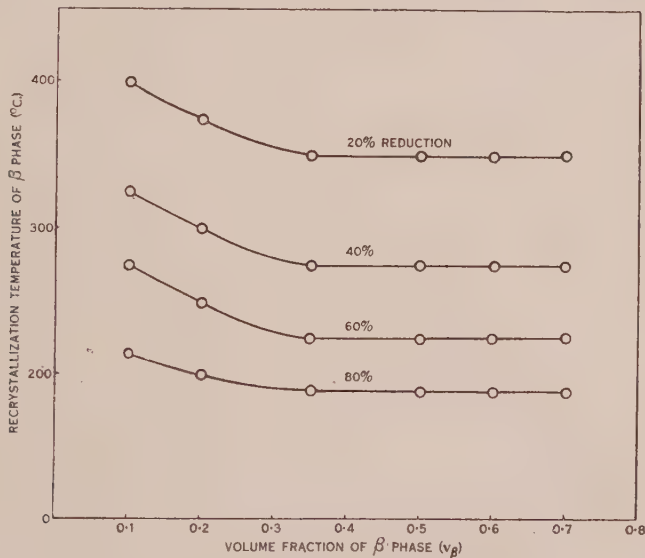


Fig. 2.—Influence of the volume fractions of the phases on recrystallization temperature of β phase after various reductions in area by wire drawing.

at various temperatures, and the recrystallization of the phases was followed by microscopic examination of longitudinal sections. A sufficient number of annealing temperatures (at 25 °C. intervals) was used to enable the degree of

recrystallization of the phases to be classified as just started, started, partly, nearly fully, and fully. The recrystallization behaviour of the phases in the various alloys could thus be compared. Although X-ray and microscopic methods give the same results, microscopic examination is faster and more reliable particularly when the size of the recrystallized grains is small. For this reason microscopic examination alone was used in these experiments.

III. RESULTS

Figures 1 and 2 show the recrystallization temperatures of the α and β phases respectively for various volume fractions of the β phase (v_β) and for various amounts of deformation. Each point in these figures represents the results of the metallographic examination of at least eight specimens. It can be seen that when v_β is greater than about 0.35 the recrystallization temperature of each phase is independent of v_β for any deformation. This indicates that the deformation of the phases is constant over this range of v_β for all total deformations used.

However, when v_β is less than 0.35 the recrystallization behaviour is different. Figure 1 shows that the deformation of the α phase is a maximum (recrystallization temperature a minimum) when $v_\beta = 0.2$. The results for 80 per cent. reduction in area are inconclusive as after heavy deformations the recrystallization temperature becomes less sensitive to changes in deformation. Figure 2 shows that the deformation of the β phase increases steadily with v_β up to approximately $v_\beta = 0.35$. The change in recrystallization temperature is less at higher deformations as would be expected from the decreased sensitivity.

The mode of recrystallization of the phases, for total deformations of up to 40 per cent. reduction in area, is the same as that described for silver-magnesium alloys (Clarebrough 1950). The hard β phase always starts to recrystallize at its boundary with the α phase. The α phase starts to recrystallize at or near its boundary with the β phase when v_β is small, but as v_β increases, nucleation in the α phase occurs further away from the α - β boundary until, when v_β exceeds approximately 0.35, nucleation in the α occurs at random. After 60 and 80 per cent. reduction in area, nucleation occurs at random in the α phase at all values of v_β , but in the β phase only when $v_\beta > 0.35$. This indicates that, whereas at the smaller reductions the deformation is inhomogeneous within the crystals of one phase and the inhomogeneity is also of a different type in the two phases, at heavier reductions the inhomogeneity is much less pronounced, probably as a result of work hardening. Typical photomicrographs illustrating the mode of recrystallization are shown in Plate 1, Figures 1-4.

IV. DISCUSSION

In a previous paper (Clarebrough 1950) it was shown that the volume fractions of the phases and their relative deformations are related by

$$p = p_\alpha + v_\beta(p_\beta - p_\alpha), \quad \dots \dots \dots (1)$$

where p = deformation per unit volume of the duplex alloy,

p_α = deformation per unit volume of the α phase,

and p_β = deformation per unit volume of the β phase.

It follows that if p_α and p_β are constant over a range of values of v_β then in this range

$$p_\alpha = p_\beta = p. \quad \dots\dots\dots (2)$$

The present results have shown that, for the alloys in the range of v_β values from 0.35 to 0.7, p_α and p_β are constant. This holds for all the deformations used. In this range, therefore, equation (2) must hold. It seems very unlikely that equation (2) would cease to apply for v_β values between 0.7 and 1.0 as the deformation is here governed by the hard β matrix. This point was not checked

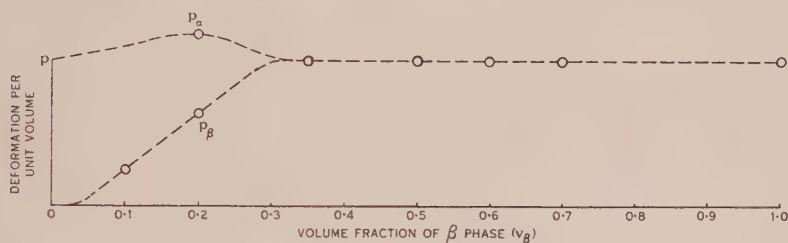


Fig. 3.—Influence of the volume fractions of the phases on their relative deformation.

experimentally as alloys in which v_β exceeds 0.7 are too brittle to deform readily. When the β phase occupies less than about one-third the volume of the alloy, p_α and p_β are no longer constant, and hence equation (2) is inapplicable.

If we now assume that in the range of $0 < v_\beta < 0.35$, p_β increases almost linearly with v_β and that equation (2) applies when $v_\beta > 0.35$, then the variations in p_α and p_β over the whole range of v_β values are shown schematically in Figure 3. A diagram of this type was put forward tentatively in a previous paper (Clarebrough 1950) and the present results are in good agreement with it, particularly in the occurrence of a maximum for p_α when v_β is near 0.2.

Thus, it can be seen that in duplex brass, as long as the volume fraction of the β phase is small, the α phase always deforms more than the β . However, once the volume fraction of the β phase exceeds approximately one-third, both the soft α and the hard β phases deform to the same extent. Duplex silver-magnesium (Clarebrough 1950) and copper-iron (Perger, unpublished data) alloys behave in the same way. It is probable then that these results apply generally to duplex alloys in which both phases are plastic, but differ considerably in yield point and work-hardening characteristics.

V. REFERENCES

- CLAREBROUGH, L. M. (1950).—*Aust. J. Sci. Res. A* **3**: 72.
 HONEYCOMBE, R. W. K., and BOAS, W. (1948).—*Aust. J. Sci. Res. A* **1**: 70.
 UNCKEL, H. (1937).—*J. Inst. Met.* **61**: 171.

EXPLANATION OF PLATE 1

Photomicrographs showing the mode of recrystallization of the phases.

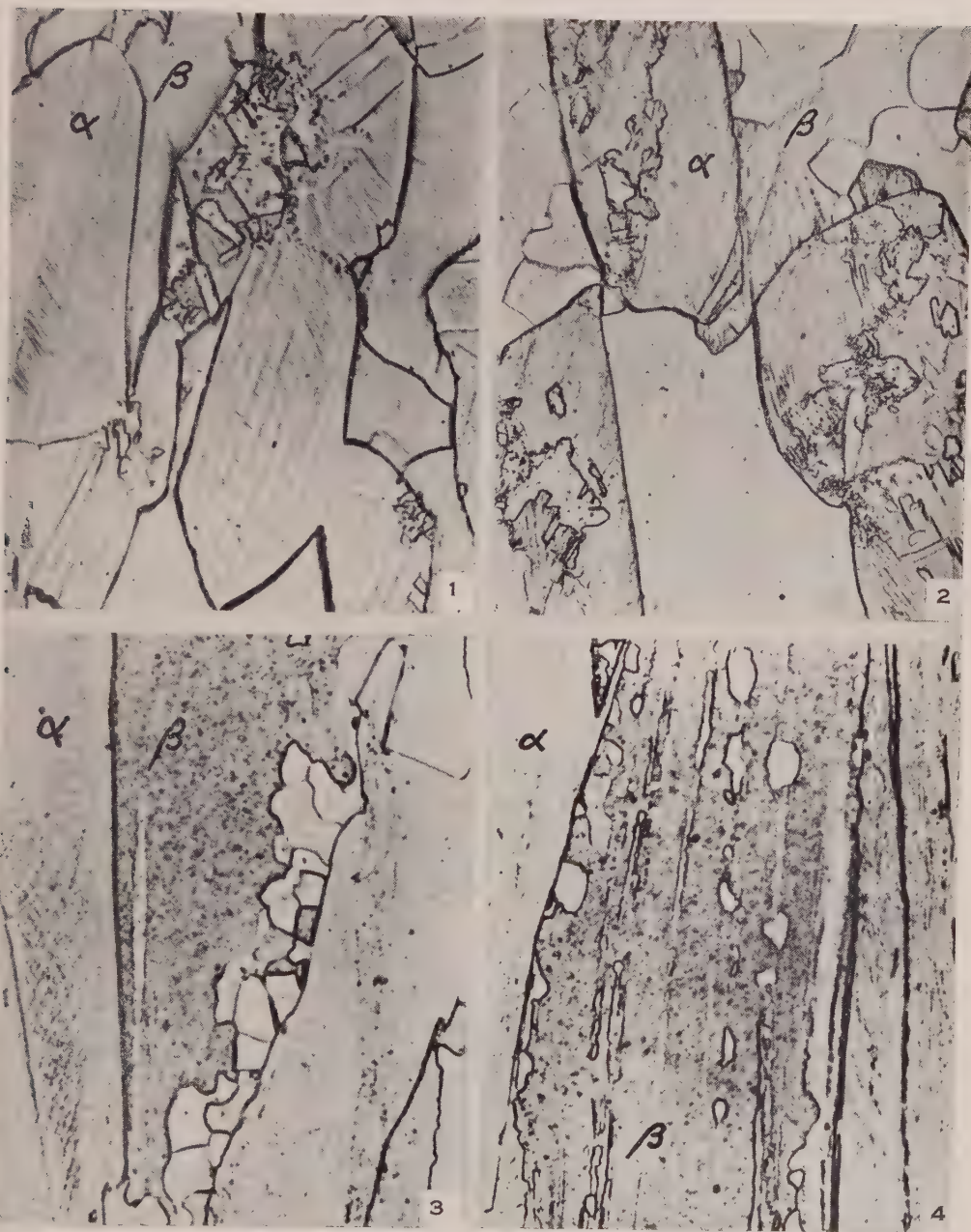
Fig. 1.— $v_{\beta}=0.35$. Reduction in area=40 per cent. Recrystallization of α phase starts at α - β boundary. $\times 520$.

Fig. 2.— $v_{\beta}=0.5$. Reduction in area=40 per cent. Recrystallization of α phase occurs at random. $\times 520$.

Fig. 3.— $v_{\beta}=0.5$. Reduction in area=40 per cent. Recrystallization of β phase starts at α - β boundary. $\times 690$.

Fig. 4.— $v_{\beta}=0.7$. Reduction in area=60 per cent. Recrystallization of β phase occurs at random. $\times 860$.

DEFORMATION OF $(\alpha + \beta)$ BRASS



THE ANISOTROPY OF ELECTRICAL RESISTIVITY OF COLD-DRAWN WIRES OF SOME CUBIC METALS AND ALLOYS

By T. BROOM* and W. K. CLOTHIER†

[Manuscript received October 16, 1951]

Summary

A special A.C. technique has been devised to measure and compare the transverse resistivities of wires. In drawn wires of several metals and alloys the resistivities measured in the longitudinal and transverse directions have been shown to differ.

In copper the longitudinal resistivity was slightly less than the transverse, but in α brass and α aluminium bronze it was the greater by about 3 per cent. This effect is similar to that previously found in rolled strip, and it is suggested that an explanation may be found in terms of lattice disturbances associated with dislocations.

I. INTRODUCTION

Recent experiments (Broom 1951) showed that anisotropy of electrical resistivity could be induced in some cubic metals and alloys by cold-rolling. Generally the resistivity in the rolling direction exceeded the transverse resistivity by 1 to 2 per cent. It was suggested that a similar and greater anisotropy should occur in drawn wires and the present experiments have confirmed this for some metals. This anisotropy is of interest as the change in resistivity due to cold-work is generally measured on wires and only in the longitudinal direction.

As the resistivity of an annealed cubic metal is isotropic, anisotropy in worked metals is conveniently investigated by finding the changes, on annealing, of resistivities measured in different directions. It is often difficult to measure resistivities directly, owing to uncertainties in measurements of specimen dimensions, and therefore fractional changes in resistance are compared.

The comparison of longitudinal resistances was carried out by conventional methods. However, in order to define and compare transverse resistances of wires, it was necessary to develop a form of contact device which would give the same current flow pattern for different specimens. Further, the resistance to be measured was small (of the order of $1\ \mu\Omega$ for a copper specimen as used in the present work), necessitating the development of a technique for measurement of very low resistances with a precision comparable to that attained in the longitudinal measurements (~ 0.1 per cent.).

II. EXPERIMENTAL

(a) Materials

The analyses of the metals selected for study are given in Table 1.

The original diameter of the rods of conductivity copper and the alloys was 0.193 in., and that of the spectrographic copper rod was 0.279 in. Before

* Division of Tribophysics, C.S.I.R.O., University of Melbourne.

† Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

drawing, all the rods were annealed in evacuated Pyrex tubes at 580 °C. for half an hour and allowed to cool in air. The rods were all cold-drawn to 0.0725 in. diameter wire using wolfram carbide dies lubricated with stearic acid. The reduction in cross section per die was approximately 20 per cent.

TABLE I
ANALYSES OF MATERIALS USED

Spectrographic copper	0.0003% Ag, Ni; <0.0004% Pb
Conductivity copper	<0.01% Ag, Cd, Pb, Si; \leq 0.01% Mg, Fe
α Brass	20.21% Zn; 0.02% Fe; 0.01% Pb, Sn
α Aluminium bronze	5.67% Al; 0.28% Zn; 0.1% Fe

For the determination of transverse resistances, specimens 0.500 in. long were prepared from the wires. Their ends were ground flat and the resulting slight burrs removed by hand. Some of these specimens were then annealed under the same conditions as described above.

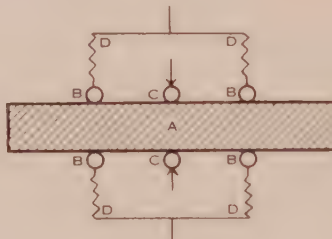


Fig. 1.—Form of contact device used in preliminary work with unamalgamated specimens (diagrammatic).

(b) Longitudinal Resistance Measurements

A conventional Kelvin double bridge was used to measure the resistance at 20 °C. of 20 cm. lengths of wire. After the drawn wires had been measured, they were annealed (together with the transverse specimens mentioned above) and remeasured.

(c) Device for Measurement of Transverse Resistance

In order to measure the extremely low values of transverse resistance it was essential to use a four-terminal arrangement of contacts. The form of the contacts had to be such that specimens could be interchanged without altering the geometrical relationship of current and potential contacts. Several different forms of contact device were tested in an attempt to fulfil these conditions.

The first arrangement, shown diagrammatically in Figure 1, consisted of a set of six hard copper wires bearing on a specimen *A*, the points of contact being in a diametral plane of the specimen, three symmetrically disposed on each side. The outer pairs *B* were connected as current leads with high resistances *D* in series with the wires to eliminate variations in current flow pattern due to changes in contact resistance. The central pair of wires *C* acted as potential contacts.

Readings obtained using this device were not very reproducible, presumably owing to relative displacements of the contact wires, but they showed that the longitudinal resistivity was higher than the transverse for both the alloys. However, as discrete contacts on the specimen obviously lead to a complex form

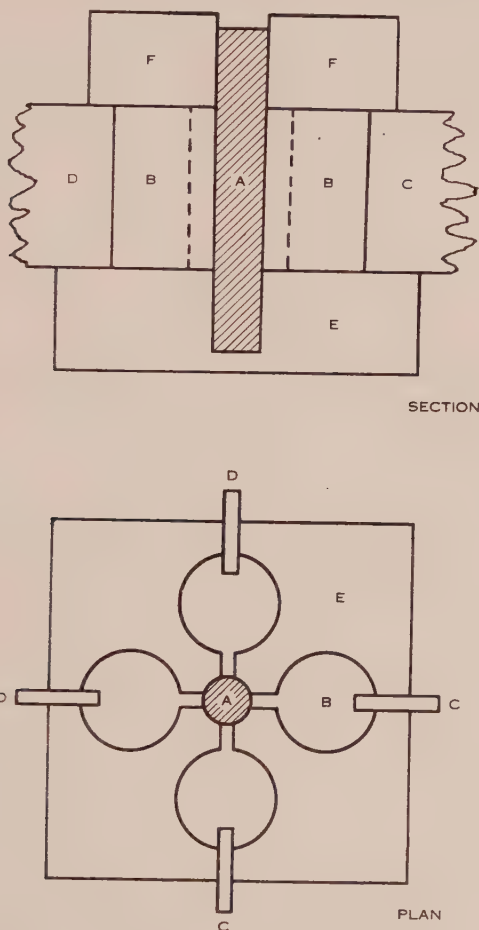


Fig. 2.—Four-contact device used in the measurement of transverse resistance of wires.

of current flow in which the longitudinal resistivity plays an uncertain part, it was desirable to develop a form of resistor in which the current leads were more extended.

The device finally used is shown in Figure 2. In a Perspex block *E*, four slots 0.020 in. wide and 0.25 in. deep communicated with the central hole 0.0735 in. diameter, 0.375 in. deep, in which was placed an amalgamated specimen *A* (0.0725 in. diameter; 0.500 in. long). The slots were kept filled with mercury from the reservoirs *B*, the level of mercury being kept constant by use of the Perspex "lid" *F*, which fitted over the specimen. Copper strips

provided electrical connections with the reservoirs such that an adjacent pair *C* were current leads and the opposite pair *D* potential leads. To ensure uniform current flow to the specimen the strips were made relatively long and the ends projecting into the reservoirs were amalgamated.

For each set of determinations freshly amalgamated specimens were prepared by dipping in mercurous nitrate solution followed by light rubbing upon filter paper. Using the circuit described below, four readings were taken for each specimen, the specimen being rotated 90° between readings. Annealed and drawn specimens of any one metal were measured alternately to reduce errors

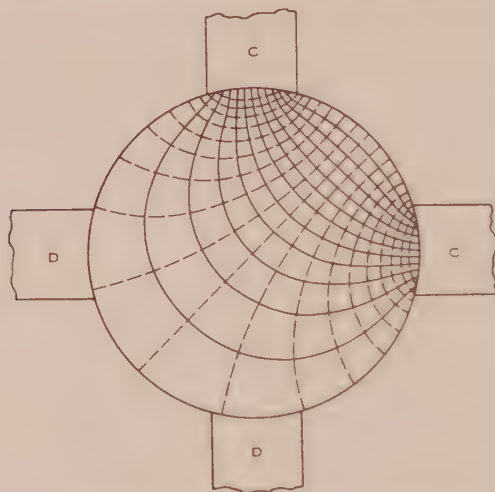


Fig. 3.—Current flow lines (full) and equipotential lines (broken) across a specimen in the apparatus shown in Figure 2.

due to temperature or other changes. Three or five such pairs of specimens were measured for each metal. The mean temperature during a set of measurements was $20.0 \pm 0.5^\circ \text{C}$.

Separate experiments showed that amalgamation had a negligible effect on the longitudinal resistivity of wires so that it is unlikely that the use of mercury contacts gave rise to any serious errors. Further, no changes of resistance values with time were noted, so that mercury attack was evidently very limited.

(d) *Measurement of Transverse Resistance*

The electrical resistance value determined in the measurement was that defined by the ratio of the potential difference between the potential leads *D*, Figure 2, to the current flowing in the current leads *C*. The form of the current flow in the specimens was determined electrically by means of a model made from thin sheet brass. The flow pattern is shown in Figure 3, where the current flow lines are shown in full and the equipotentials in broken lines.

Using the mercury contact device described above, the resistance value for copper was approximately $0.6 \mu\Omega$. It was not possible to use large measuring

currents as no special provision was made for cooling the specimen. In the present work the current was limited to 1 A. as it was found that this did not produce any significant rise in temperature of the mercury, in which the greater proportion of energy dissipation occurred owing to its high resistivity relative to copper.

Since the accuracy sought in the electrical measurements was 0.1 per cent., the minimum potential to be detected was about 6×10^{-10} V. This is beyond the limit that may be reached by conventional direct current methods of measurement employing the D'Arsonval galvanometer, partly because of the inherent inefficiency of energy transfer using this type of galvanometer in circuits of very low resistance and partly because of the instability introduced into the measurement by parasitic thermal e.m.f.'s in the galvanometer and specimen circuits.

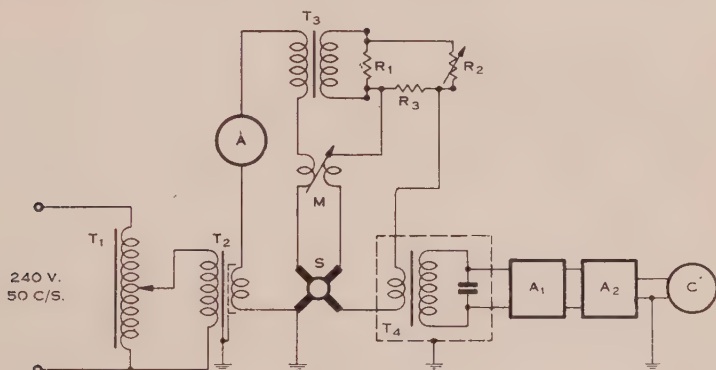


Fig. 4.—Circuit for the measurement of transverse resistance of wires.

Alternating current methods have several important advantages over D.C. methods in this class of measurement:

- (i) Parasitic thermal e.m.f.'s are of no significance.
- (ii) Advantage may be taken of the low resistance of the potential circuit by the use of a step-up transformer of high ratio.
- (iii) The measuring circuit can be arranged in a bridge form in which the conditions for balance are independent of the resistance of the current leads.

Accordingly, an alternating current method was employed using the circuit shown in Figure 4.

Current at mains frequency (50 c/s.) was supplied to the specimen *S* from a variable-ratio transformer *T*₁ and isolating transformer *T*₂. The specimen current flowed through an ammeter *A* and the primary windings of a current transformer *T*₃ and the mutual inductor *M*. The potentials from the secondary circuit of *M* and from *R*₃ in the secondary circuit of *T*₃ were adjusted so that their sum was equal to the potential difference between the potential terminals of the specimen. Equality was indicated by a balance detector which comprised a tuned step-up transformer *T*₄, a battery-operated preamplifier *A*₁, a tuned mains-operated amplifier *A*₂, and a cathode-ray oscillograph *C*.

The mutual inductor M was necessary in order to compensate for unavoidable mutual inductance between the current and potential circuits, particularly within the four-terminal contact device. However, its value did not enter into the resistance measurement.

It can readily be shown that the specimen resistance R is given by

$$R = \frac{R_1 R_3}{N(R_1 + R_2 + R_3)},$$

where N is the ratio of primary to secondary current in the transformer T_3 , and R_1 , R_2 , and R_3 are the numerical values of the corresponding resistors in Figure 4. In the measurements N was equal to 1, R_1 was 0.01 ohm, and R_3 0.1 ohm. Since R_2 was always greater than 200 ohm, to a sufficient degree of accuracy

$$R = \frac{R_1 R_3}{R_2},$$

and since only relative values of resistance were required, a knowledge of the exact values of R_1 and R_3 was unnecessary.

Further details of the electrical circuit are given in Appendix I.

III. RESULTS

Owing to the form of current flow and the relatively wide potential contacts in the four-terminal resistor, it was necessary to calibrate the apparatus using specimens of known transverse resistivity. Measurements on annealed specimens of various metals whose transverse resistivities are equal to the (known) longitudinal resistivities, showed that within experimental error, the resistances measured were directly proportional to resistivities. Therefore, it was possible to equate the fractional change of resistivity on annealing, $(\rho_D - \rho_A)/\rho_A$ (or $\Delta\rho/\rho$), to the fractional change of resistance, $(R_D - R_A)/R_A$, where the subscripts D and A refer to deformed and annealed material respectively.

Two sets of experiments (transverse measurements) with slightly different arrangements were carried out using conductivity copper and the alloys, and also a further set in which only the conductivity and spectrographic coppers were used.

Table 2 gives values of $\Delta\rho/\rho$ for both longitudinal and transverse directions. In the case of the transverse results, standard errors are given together with the number of specimen pairs (annealed and deformed) measured in each set of determinations. For the longitudinal results, maximum experimental errors are given which for the limited number of determinations in the present work are approximately equal to standard errors.

The last column of Table 2 gives the anisotropy of resistivity of the drawn wires calculated as

$$\frac{(1 + \Delta\rho/\rho)_L - 1}{(1 + \Delta\rho/\rho)_T} = \frac{(\rho_D/\rho_A)_L - 1}{(\rho_D/\rho_A)_T} = (\rho_L/\rho_T)_D - 1,$$

where the subscripts L and T refer to the longitudinal and transverse directions. This definition of anisotropy differs from that given in the previous paper (Broom 1951).

TABLE 2
ANISOTROPY OF INCREASE IN RESISTIVITY ON WIRE-DRAWING

Metal	Resistivity at 20 °C. (ohm cm. $\times 10^{-6}$) (annealed)	Deformation $\ln (A_0/A)$	Fractional Increase at 20 °C. $\Delta\rho/\rho \times 10^2$		Anisotropy of Resistivity ($\rho_L/\rho_T - 1$) $\times 10^2$
			Longitudinal	Transverse	
Spectrographic copper	1.69	2.70	2.3 ± 0.2	2.9 ± 0.2 (5)*	-0.6 ± 0.4
Conductivity copper	1.72	1.96	1.8 ± 0.2	2.7 ± 0.4 (5)	-0.9 ± 0.6
				2.8 ± 0.5 (3)	-1.0 ± 0.7
				2.3 ± 0.6 (3)	-0.5 ± 0.8
α Brass ..	5.47	1.96	18.4 ± 0.1	14.9 ± 0.3 (3)	$+3.1 \pm 0.3$
				14.8 ± 0.2 (3)	$+3.1 \pm 0.3$
α Aluminium bronze	11.13	1.96	27.4 ± 0.1	24.4 ± 0.2 (3)	$+2.4 \pm 0.3$
				23.9 ± 0.4 (3)	$+2.8 \pm 0.4$

* Figures in brackets refer to the number of specimen pairs (annealed and drawn) used in determining $\Delta\rho/\rho$ (transverse).

IV. DISCUSSION

The results indicate that the transverse resistivities of all the wires used differed from their longitudinal resistivities. It should be noted that the current flow in the transverse experiments had a longitudinal component as the current leads only made contact over the central quarter inch of the specimens. The result of this is that the magnitude of the measured anisotropy is smaller than that of the actual anisotropy. The values of anisotropy may now be compared with those found for cold-rolled strip. For copper, the resistivity is isotropic (Broom 1952) or only slightly anisotropic (Broom 1951), the results for strip and wire being very similar. In the alloy wires the anisotropy is much greater than that produced by cold-rolling similar single phase alloys.

In the earlier paper it was suggested that anisotropy of resistivity is due to anisotropic scattering of conduction electrons by oriented dislocations. A recent theory of the increase of resistivity due to cold-work (Mackenzie and Sondheimer 1950) predicts maximum scattering in the direction of slip. In the fibre structures developed by wire-drawing, the operative slip planes are less inclined to the wire axis than they are to the direction of rolling in strip, and hence the anisotropy in a wire should be greater than in a strip. This has indeed been found. However, it seems unlikely that both positive and negative values of anisotropy can be accounted for in terms of the previous suggestion, although an explanation might be found in terms of the possible consequences of dislocations, such as the creation of stacking faults and the segregation of solute atoms.

It may be noted that if the temperature coefficient of resistivity of drawn wire is isotropic, as is the case in rolled strip and in hexagonal single crystals, some deviations from Matthiessen's rule can be expected.

V. ACKNOWLEDGMENTS

The authors wish to thank Mr. H. Bairnsfather and Mr. J. F. Nicholas for assistance in the experimental work.

VI. REFERENCES

- BROOM, T. (1951).—*Phil. Mag.* **42**: 56.
 BROOM, T. (1952).—*Aust. J. Sci. Res. A* **5**: 128.
 MACKENZIE, J. K., and SONDHEIMER, E. H. (1950).—*Phys. Rev.* **77**: 264.

APPENDIX I

Details of Electrical Components and Measuring Technique

In the circuit shown in Figure 4, T_3 was a conventional current transformer with primary and secondary both rated at 5 A. Its ratio error and phase angle were negligible, but a transformer with much larger errors could have been used since the primary current and secondary burden were practically constant.

The tuned detector transformer T_4 , which had a step-up ratio of 3500:1 was of great importance in the technique. Its purpose was to increase the signal voltage so that it could be observed in the presence of thermal and flicker noise from the amplifier A_1 . The transformer had a core of thin interleaved Mumetal and its secondary was tuned with mica capacitors to the mains frequency of 50 c/s. By this means it was found possible to obtain a high step-up ratio and at the same time an input impedance of sufficient value to match the resistance of the potential circuit. The transformer was of "core" type construction with primary and secondary equally divided between the two limbs, this form giving low interference from stray magnetic fields. It was enclosed within a heavy gauge Mumetal shield.

A battery preamplifier A_1 was used to increase the signal voltage above the hum level of the mains-operated amplifier A_2 . The tuned amplifier A_1 and cathode-ray oscillograph C were conventional.

The mutual inductor M was of very simple construction, its primary being wound on an insulating tube and its secondary on an insulating rod sliding within the primary. The secondary winding was astatic in form in order to avoid interference from stray magnetic fields.

Since extremely small potential differences were significant in the measurements, it was essential to reduce the loop area of the potential circuit to a minimum, not only to avoid unwanted electromagnetic coupling with the current supply circuit, but also to avoid induction from stray alternating fields of mains frequency. Therefore the wiring of the potential circuit was carried out in twisted pair and R_3 was in the form of a bifilar loop of resistance wire. R_1 and R_3 were four-terminal resistors and R_2 was a low-inductance resistance box.

Stray induction was further reduced by locating the transformers T_1 , T_2 , and T_3 and the mains-operated units of the detector circuit well away from the potential circuit.

It was necessary to avoid multiple earthing in the detector circuit, i.e. earthing at two or more points already connected by low impedance paths. Multiple earthing may lead to circulating currents induced by stray fields in the low impedance loops so formed and thus to spurious potential differences in the interconnecting leads of the detector units.

ON THE ANISOTROPY OF ELECTRICAL RESISTIVITY OF DEFORMED CUBIC METALS AND ALLOYS

By T. BROOM*

[*Manuscript received October 16, 1951*]

Summary

The change of resistivity on annealing cold-rolled metals has been found to depend on the direction of measurement for copper-nickel, steel, α brass, and sterling silver, while no such variation was observed for aluminium, copper, or nickel. An explanation of anisotropy both in strip and in wire is advanced in terms of lattice disturbances such as dislocations, vacant sites, and faults, some of which can give rise to anisotropy of resistivity when there is a preferred orientation of crystals.

Results are presented showing the influences of deformation, recovery, and a second phase on the anisotropy developed in brasses.

I. INTRODUCTION

Recent experiments (Broom 1951) have shown that anisotropy of electrical resistivity may be developed in cubic alloys by cold-rolling. The present investigation was carried out to resolve some uncertainties in the former work, namely, whether anisotropy is produced in copper, and whether the anisotropy is influenced by the presence of a second phase. Some experiments have also been carried out on nickel and 75.25 copper-nickel strip, and a limited investigation has been made of the effects of degree of deformation and of recovery on some copper-zinc alloys.

II. EXPERIMENTAL

The materials studied are listed in Table 1 together with their analyses and the annealing treatments adopted throughout the present work. The equilibrium diagram of the copper-zinc system due to Raynor (1946) was used to determine the phase relationships of the brasses given in the fourth column of Table 1. Salt baths were used for annealing the original strips (before rolling).

Most experiments were carried out using strips initially 2.5 mm. thick and about 60 mm. wide, which were rolled to 0.125–0.150 mm. thickness. The effect of the degree of deformation was studied using strips of different original thickness rolled to the same final thickness.

As before (Broom 1951), the principle of the investigation was to determine statistically whether the change of resistivity on annealing, relative to the resistivity of the annealed material, was significantly different for specimens cut at 0 and 90° to the rolling direction.

Major sources of experimental error in the earlier work were non-parallelism of the sides of the 2 mm. wide specimens, and inaccurate location of the current

* Division of Tribophysics, C.S.I.R.O., University of Melbourne.

and potential contacts which should lie along the centre-line of a specimen. To reduce these errors, the techniques of specimen preparation and measurement were modified. A new apparatus (Plate 1, Fig. 1) was used in conjunction with a guillotine for cutting the specimens for the present work. The cold-rolled strip, accurately trimmed to width by a previous operation, was advanced towards the blade of the guillotine by means of the lathe slide. Specimens of

TABLE I
MATERIALS USED

Material	Analysis	Annealing Treatment	Structure	Assumed Density (g. cm. ³)
Copper	<0.01% Ag, Pb, Si; <0.01% Fe	500 °C., 1 hr., quench	Single phase	8.94
Nickel	0.024% Co; 0.14% Si; 0.07% Mg; 0.05% C; <0.05% Mn, Cr, Cu; <0.01% Fe, Al	600 °C., $\frac{1}{2}$ hr., quench	Single phase	8.89
75 25 Copper-nickel	74.8% Cu; 0.1% Al; 0.02% C; 0.07% Mn; <0.05% Pb, Sn, Zn, Fe	600 °C., $\frac{1}{2}$ hr., quench	Single phase	8.94
70 30 Brass	70.2% Cu; 0.01% Sn, Ni; <0.01% Pb, Fe	500 °C., 1 hr., quench	α , Single phase	8.53
64 36 Brass	64.4% Cu; 0.01% Sn, Ni; <0.01% Pb, Fe	600 °C., $\frac{1}{2}$ hr., quench	α , Single phase supercooled at R.T.	8.47
60 40 Brass	60.0% Cu; 0.05-0.10% Ni; <0.01% Pb, Fe	450 °C., 2 hr., quench	$\alpha+\beta$; 15% β	8.39
		700 °C., $\frac{1}{2}$ hr., quench	$\alpha+\beta$; 77% β	

equal width were obtained using a 1.00 mm. thick steel spacer to define the distance between the fixed dial gauge and a screw moving in an arm attached to the slide. The screw was adjusted to give a gauge reading of zero after a cut had been made and the spacer removed. The slide was then advanced, the spacer re-inserted, and the dial gauge reading brought to zero by careful adjustment of the slide. A cut was then made, the strip having been advanced by the thickness of the spacer. This procedure yielded parallel-sided specimens with weights generally constant to within 2 per cent.

The new four-point contact device is shown in Plate 1, Figure 2. The potential contacts were 30 mm. apart compared with 20 mm. in the previous work, and all soldered connections between the contacts and leads to the Kelvin double bridge were immersed in the constant temperature bath, so reducing stray thermal e.m.f.'s.

Ten specimens were cut parallel to the rolling direction and 10 in the transverse direction. The resistance, weight, and length of all specimens were measured and the resistivities calculated using assumed values for the densities (Table 1). Some of the specimens were given a "recovery" treatment by

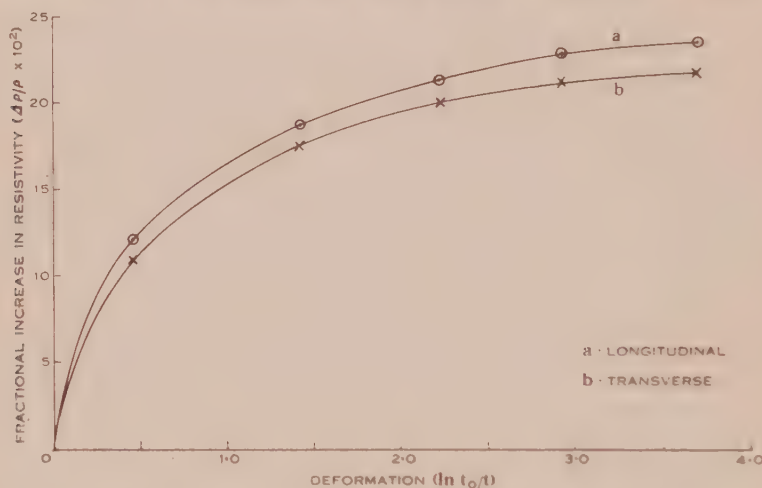


Fig. 1.—Increase in electrical resistivity of 70/30 brass on cold-rolling.

heating at 150 °C. for 12 hours in air and their resistances measured before the final annealing treatment and remeasurement. Annealing was carried out in evacuated tubes of Pyrex or silica (for the 700 °C. treatment).

III. RESULTS

In Table 2 are given values of the fractional changes of resistivity, $\Delta\rho/\rho$, on annealing (and their standard errors) for the longitudinal and transverse directions. The values in italics refer to fractional changes on annealing specimens which had been given the "recovery" treatment. The anisotropy of resistivity is given as $\rho_L/\rho_T = 1$, where the suffixes *L* and *T* refer to the longitudinal and transverse directions. This definition differs from that given by Broom (1951). The deformations, calculated as logarithmic strains, $\ln t_0/t$ (where *t* is the strip thickness), to which the values of anisotropy refer, are also given in Table 2 together with the mean resistivities of annealed longitudinal specimens.

The effect of deformation on the anisotropy developed in 70/30 brass is shown in Figure 1, where the fractional changes in resistivity in the longitudinal and transverse directions are plotted against deformation expressed as logarithmic strain.

TABLE 2
ANISOTROPY OF INCREASE IN RESISTIVITY ON COLD-ROLLING

Metal	Resistivity at 20 °C. (ohm cm. 10 ⁻⁶) (annealed)	Deformation ln (<i>t</i> ₀ / <i>t</i>)	Fractional Change in Resistivity at 20 °C. $\Delta \rho / \rho \times 10^2$			Anisotropy of Resistivity ($\rho_L / \rho_T - 1$) $\times 10^2$
			Long.	Trans.	Error	
Copper	1.684	2.919	3.20	3.27	± 0.14	-0.07 ± 0.27
Nickel	9.939	2.901	3.44	3.39	± 0.11	$+0.05 \pm 0.21$
75/25 Copper-nickel	30.56	2.226	3.34	4.03	± 0.16	-0.66 ± 0.30
70/30 Brass (α)	6.164	2.919	22.97	21.16	± 0.19	$+1.49 \pm 0.32$
	6.167	3.698	23.49	21.77	± 0.14	$+1.41 \pm 0.23$
		2.919*	16.61	15.39	± 0.19	$+1.06 \pm 0.33$
64/36 Brass (α)	6.493	2.919	24.37	23.25	± 0.18	$+0.91 \pm 0.29$
	6.413	3.714	26.71	24.92	± 0.15	$+1.43 \pm 0.25$
		2.919	14.95	14.50	± 0.18	$+0.39 \pm 0.32$
60/40 Brass (15% β)	6.865	2.883	21.05	19.89	± 0.38	$+0.97 \pm 0.64$
	6.971	3.634	20.13	18.60	± 0.26	$+1.29 \pm 0.44$
		2.883	7.16	6.59	± 0.38	$+0.53 \pm 0.72$
60/40 Brass (77% β)	6.491	2.901	35.94	33.08	± 0.21	$+2.15 \pm 0.32$
	6.431	3.650	40.55	36.04	± 0.16	$+3.32 \pm 0.23$
		2.901	20.65	18.63	± 0.21	$+1.70 \pm 0.36$

* Values in italics refer to changes of resistivity on fully annealing specimens after a "recovery" treatment at 150 °C. for 12 hours.

IV. DISCUSSION

In a previous paper (Broom 1951) an explanation of anisotropy of resistivity in rolled cubic metals was put forward. In view of the new data presented in this paper for rolled strip, and elsewhere (Broom and Clothier 1952) for drawn wire, the previous explanation must be re-examined.

The metals investigated can be classified according to whether the longitudinal resistivity is larger than, equal to, or less than the transverse resistivity :

- (i) ($\rho_L / \rho_T - 1$) > 0 : Steel, brass, sterling silver strip.
Brass, aluminium bronze wire.
- (ii) ($\rho_L / \rho_T - 1$) $= 0$: Copper, nickel, aluminium strip.
- (iii) ($\rho_L / \rho_T - 1$) < 0 : Copper,* copper-nickel strip.
Copper wire (spectrographic and commercial purities).

* Broom (1951).

Inspection of Table 2 shows that there is no relation between the anisotropy and either the fractional or the absolute increase in resistivity on deformation. For instance, although copper-nickel shows a relatively small increase in resistivity, the anisotropy is of the same magnitude as that in those alloys (such as α brass) which have a large increase in resistivity.

In the earlier paper it was assumed that the effect of deformation on electrical resistivity was wholly due to dislocations. Owing to the anisotropy of their strain fields such lattice disturbances will scatter the conduction electrons anisotropically and any non-random array of dislocations would be expected to give rise to anisotropy of electrical resistivity. The production of a deformation texture will result in preferred orientations of dislocations, so that anisotropy of resistivity is possible in deformed metals. The magnitude of the effect may be calculated if assumptions are made about the anisotropy of individual dislocations* and about their distribution among the slip planes of various ideal deformation textures. In this way, some qualitative agreement with experiment may be obtained, but it is difficult to account both for positive anisotropy in α brass and negative anisotropy in copper-nickel, since their rolling textures are the same (Göler and Sachs 1929), and there is no obvious reason why the dislocations should be distributed differently among the slip planes.

However, in real metals with a close-packed structure, a dislocation consists of two half-dislocations separated by a region in which the stacking sequence of the lattice is faulty (Heidenreich and Shockley 1948). The equilibrium separation of these half-dislocations will vary from metal to metal, as it depends on the difference between the energy of the lattice with normal stacking and that of the lattice with faulty stacking. Therefore, although individual half-dislocations in various metals may have the same anisotropy of resistance, the extent of the faulted regions may differ and thus cause different anisotropy of resistivity. For reasons which will be discussed elsewhere, it is likely that such faults are more prominent in α brass than in copper-nickel.

Further, vacant lattice sites or "holes" produced by deformation may be important in determining resistivity changes. Individual holes scatter the conduction electrons isotropically, but an anisotropic array of single holes may give rise to anisotropy of resistivity. This effect will probably be small if the mean spacing between holes is greater than the mean free path of the conduction electrons. In considering aggregates of holes, Seitz (1950) has pointed out that a disk-shaped aggregate would be the most stable and that above a certain size it would break down into a ring dislocation enclosing a volume of faulty packing. Anisotropy of resistivity could obviously be expected if non-randomly oriented arrays of either disk-shaped aggregates or ring dislocations exist.

Thus, although it is not possible to account in detail for the observed anisotropy in various metals and alloys, it seems that a satisfactory explanation may be given if the structure and form of dislocations in real metals are considered.

* Mackenzie and Sondheimer's (1950) theory suggests that the dislocation resistance in the direction of slip should be twice that in a direction perpendicular to the slip plane.

It will be noted from Figure 1 that anisotropy of about 1 per cent. was found in α brass rolled to a reduction in thickness of only 37 per cent. However, the results of Figure 1 were obtained with strips which had all been rolled to 0.125 mm. thickness from strips of different thicknesses, and no precautions were taken to ensure that the original annealed strips had comparable recrystallization textures. As an existing texture influences the development of another texture on deformation and the anisotropy of resistivity is very sensitive to the orientation of the crystals, it is not certain whether the values of anisotropy observed in the brass are strictly comparable.

The persistence of anisotropy following recovery in the 70/30 brass suggests (in the absence of any change of crystal orientation at 150 °C.) that recovery causes only a slight adjustment in the number or size of those lattice disturbances which cause increased resistivity. The change in resistivity with recovery is small and of the same order of magnitude as the change in anisotropy.

The results obtained with the 64/36 brass agree well with those obtained previously with a brass of very similar composition (Broom 1951). The anisotropy developed is of the same order as for the 70/30 brass. However, in this case the anisotropy retained on recovery is much less and the resistivity decrease is larger. An explanation seems possible on the basis of a hypothesis advanced by Clarebrough (1950*a*) to account for higher recrystallization temperatures in duplex alloys when concurrent precipitation occurs. It was suggested that dislocations move to the surfaces of precipitating crystals, thus decreasing the energy per unit volume. In 64/36 brass, as heat-treated and worked, precipitation of β could be expected on heating at 150 °C., and this would result in a larger decrease in the number of lattice disturbances (and hence resistivity) than in the 70/30 brass which remained single phase.

The anisotropy of the two-phase brasses will be influenced by various factors such as the geometrical distribution of the phases (Chang and Guenot 1950), their relative resistivities and deformations (Honeycombe and Boas 1948; Clarebrough 1950*b*), and by the degree of anisotropy developed in each phase. Therefore, it is not proposed to discuss the present results, except to point out that in duplex brass a small volume proportion of β phase does not greatly affect either the increase in resistivity or the anisotropy.

V. ACKNOWLEDGMENTS

The author's thanks are due to Dr. W. Boas for his advice and encouragement and to Mr. E. D. Hondros for assistance in the experimental work. Grateful acknowledgment is made to Austral Bronze Co. Pty. Ltd. for the gift of the copper and the brasses, and to the Royal Mint, Melbourne, for the copper-nickel alloy.

VI. REFERENCES

- BROOM, T. (1951).—*Phil. Mag.* **42**: 56.
 BROOM, T., and CLOTHIER, W. K. (1952).—*Aust. J. Sci. Res. A* **5**: 119.
 CLAREBROUGH, L. M. (1950*a*).—*Nature* **165**: 39.
 CLAREBROUGH, L. M. (1950*b*).—*Aust. J. Sci. Res. A* **3**: 72.

- CHANG, P. L., and GUENOT, R. (1950).—*J. Iron Steel Inst.* **165** : 166.
- GÖLER, F. v., and SACHS, G. (1929).—*Z. Phys.* **56** : 477.
- HEIDENREICH, R. D., and SHOCKLEY, W. (1948).—Rep. Conf. on Strength of Solids, Phys. Soc. London, p. 57.
- HONEYCOMBE, R. W. K., and BOAS, W. (1948).—*Aust. J. Sci. Res. A* **1** : 70.
- MACKENZIE, J. K., and SONDHEIMER, E. H. (1950).—*Phys. Rev.* **77** : 264.
- RAYNOR, G. V. (1946).—Annotated Equilibrium Diagrams, No. 3. Institute of Metals, London.
- SEITZ, F. (1950).—*Phys. Rev.* **79** : 890.

ON ANISOTROPY OF ELECTRICAL RESISTIVITY OF DEFORMED METALS

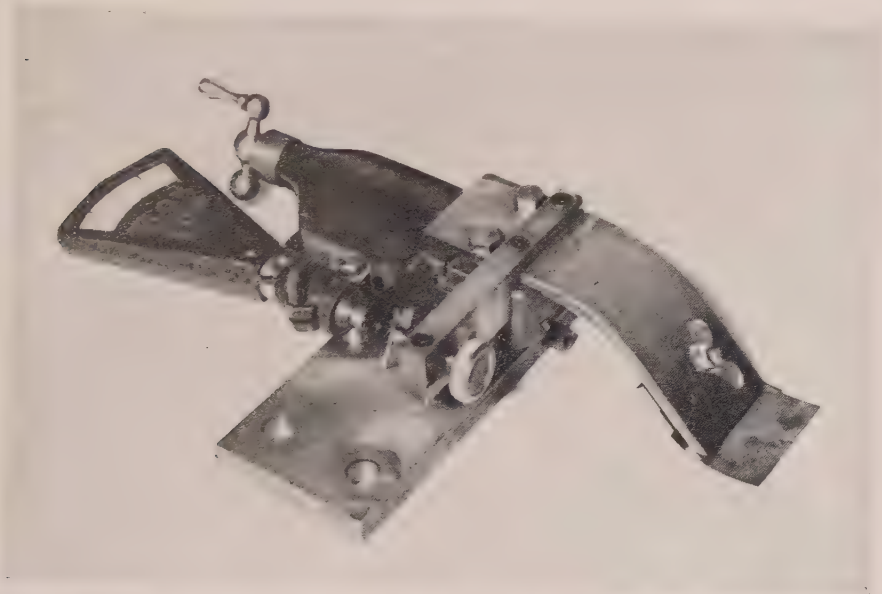


Fig. 1.—Guillotine attachment used in cutting specimens from rolled strip.

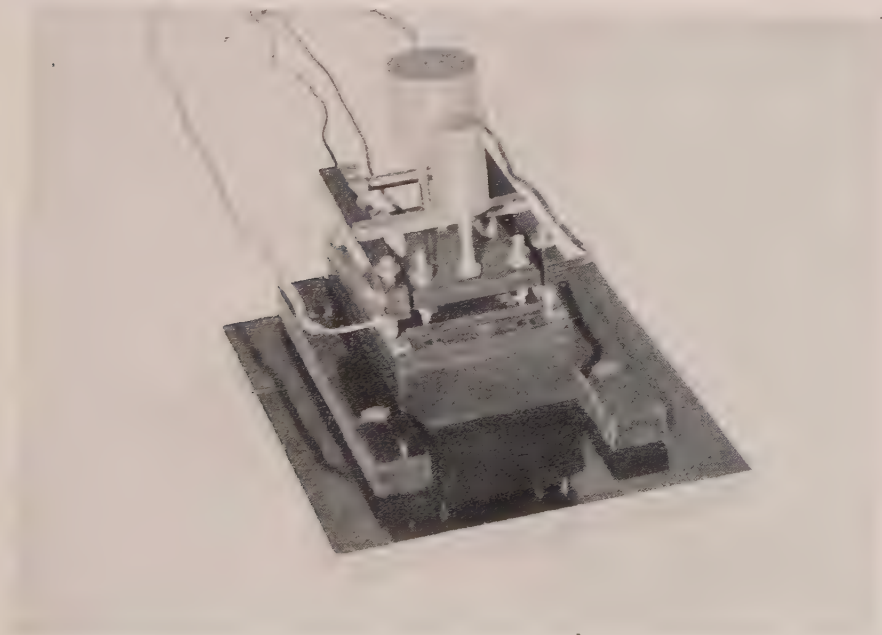


Fig. 2.—Four-point contact device for resistance measurements.

THE DIELECTRIC PROPERTIES OF SYSTEMS CONTAINING STRAIGHT POLAR CHAINS

By R. A. SACK*

[*Manuscript received July 13, 1951*]

Summary

A mathematical derivation is given of the dielectric properties of systems containing straight polar chains such that to each value of the total moment of a chain there corresponds only one arrangement of its dipoles. If the moments of the individual dipoles and the probability of an elementary transition are fixed, both the total dielectric loss and the effective relaxation time of the system increase in proportion to the square of the number of states of each chain. These conclusions are not valid for kinked chains and apply only qualitatively if the chains are branched.

The theory provides an explanation for the high dielectric losses at low frequencies observed in many solids containing hydroxyl groups. It can further explain the low frequency absorption found in ionic crystals containing lattice imperfections; in this interpretation the theory is related to Jaffé's theory of conductivity in polarizable media.

I. INTRODUCTION

It has been found experimentally that many dipolar solids containing hydroxyl groups show a dielectric absorption which is considerably larger and occurs at much lower frequencies than that of substances of similar chemical constitution, but with other polar groups. This has been found to be the case in certain modifications of primary and secondary long-chain alcohols (Baker and Smyth 1938; Hoffman and Smyth 1949; Meakins and Mulley 1951; Meakins and Sack 1951; Meakins and Welsh 1951), in the glycerides of fatty acids (Crowe and Smyth 1950), in ice (see, for example, Lamb 1946), and in various other solids and viscous liquids. It has been suggested in a previous paper (Meakins and Sack 1951) that this high loss could be explained if the hydroxyl groups were arranged in chains by means of hydrogen bonds; such a chain could alter or reverse its dipole moment by rotation of the individual hydroxyl groups. The mechanism is such that the more links the chain contains the greater will be the proportion of dipoles pointing in the direction of an applied field, but the greater will be the number of elementary steps necessary to reach a state of equilibrium after any change of field strength. This means that with increasing length of the chain both the polarizability and the effective relaxation time increase.

In the present paper the mathematical treatment of this mechanism is developed. The same formalism is applicable to other mechanisms, in particular to the motion of lattice imperfections over a limited range within ionic crystals,

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

and could thus afford an explanation of the low frequency losses observed in various crystals by Breckenridge (1948, 1950).

For long chains the discrete nature of the elementary processes becomes unimportant, and the finite difference equations by which they are expressed can be replaced by differential equations which represent a continuous process of the nature of a diffusion mechanism. This differential formulation is similar to, though less general than one put forward by Jaffé (1933) for static fields; its application to variable fields is new.

II. MODEL AND CALCULATIONS

The mathematical model used is that of a number of straight "chains" each of which can exist in any one of n different "states" for which the effective electric moments μ_j in the direction of the chain can be expressed as

$$\mu_j = jm, \quad \dots\dots\dots (1)$$

where m is a measure of the elementary dipole and j can take up all the integral or half-integral values between $-\nu$ and $+\nu$, where

$$n = 2\nu + 1. \quad \dots\dots\dots (2)$$

In the absence of an electric field all states of a chain are assumed to be occupied with equal probability p_0 . Transitions are considered possible only between successive states, i.e. in an elementary process the suffix j can change by one unit only.

If the chain is acted upon by a constant internal field F_x along its length, the probability p_j of finding it in the j th state will be, by Boltzmann's theorem,

$$p_j = \text{const.} \exp(jmF_x/kT), \quad \dots\dots\dots (3)$$

or for sufficiently small values of F_x for which saturation effects can be neglected

$$p_j = p_0(1 + jmF_x/kT). \quad \dots\dots\dots (4)$$

The total moment M_n of each chain will be on an average

$$\begin{aligned} M_n &= \sum_{-\nu}^{\nu} j m p_j = \sum_{-\nu}^{\nu} j m p_0 (1 + jmF_x/kT) \\ &= \frac{1}{3} R \nu(\nu+1)(2\nu+1) = \frac{1}{12} R(n^3 - n), \quad \dots\dots\dots (5) \end{aligned}$$

where

$$R = m^2 p_0 F_x / kT. \quad \dots\dots\dots (6)$$

If the probabilities per unit time of a system in the j th state changing to the $(j+1)$ th and $(j-1)$ th states be denoted by w_+ and w_- respectively, the condition of statistical equilibrium demands

$$p_j w_- = p_{j+1} w_+, \quad \dots\dots\dots (7)$$

which with the use of (4) yields

$$w_+ = w_0(1 + \frac{1}{2} m F_x / kT), \quad w_- = w_0(1 - \frac{1}{2} m F_x / kT), \quad \dots\dots\dots (8)$$

where w_0 is the transition probability for an elementary step in either direction in the absence of a field.

In the case of a variable field the transition probabilities depend only on the instantaneous value of F_x , but the probabilities of occupation p_j are now

given by a set of differential equations, which express the rate of change of the number of chains in a particular state as the number entering it in unit time, minus the corresponding number leaving it.

Hence

$$dp_j/dt = w_+ p_{j-1} + w_- p_{j+1} - (w_+ + w_-) p_j, \quad (j = -\nu + 1, \dots, \nu - 1), \quad \dots (9)$$

and for the extreme states

$$dp_\nu/dt = w_+ p_{\nu-1} - w_- p_\nu, \quad dp_{-\nu}/dt = w_- p_{-\nu+1} - w_+ p_{-\nu}. \quad \dots (10)$$

Confining ourselves to the case of sinusoidally varying fields $F_x = F_{x0} e^{i\omega t}$ and putting, in analogy to (4),

$$p_j = p_0 (1 + m F_{x0} e^{i\omega t} \varphi_j / kT), \quad \dots (11)$$

we obtain for the coefficients φ_j from (9) with the use (8)

$$i\omega \varphi_j / w_0 = 2i\omega \tau_0 \varphi_j = \varphi_{j+1} - 2\varphi_j + \varphi_{j-1}, \quad \dots (12)$$

and for the boundary conditions (10)

$$2i\omega \tau_0 \varphi_\nu = 1 - \varphi_\nu + \varphi_{\nu-1}, \quad 2i\omega \tau_0 \varphi_{-\nu} = -1 - \varphi_{-\nu} + \varphi_{-\nu+1}. \quad \dots (13)$$

Here the time $\tau_0 = 1/2w_0$ represents the time of relaxation of an elementary transition.

The difference equation (12) has the general solution

$$\varphi(j) = A \sinh j\alpha + B \cosh j\alpha, \quad \dots (14)$$

where the parameter α is determined by the relation

$$\cosh \alpha = 1 + i\omega \tau_0, \quad \sinh^2 \frac{1}{2}\alpha = \frac{1}{2}i\omega \tau_0. \quad \dots (15)$$

The coefficients A and B are found from (13) as

$$B = 0, \quad 1/A = (1 + 2i\omega \tau_0) \sinh \nu\alpha - \sinh (\nu - 1)\alpha = \sinh (\nu + 1)\alpha - \sinh \nu\alpha, \quad \dots (16)$$

The average moment M_n of the chain is given similar to (5) as

$$M_n = \sum_{-\nu}^{+\nu} j m p_j = R \sum_{-\nu}^{+\nu} j \varphi_j, \quad \dots (17)$$

where the R defined in (6) now refers to the alternating field. The sum in the expression (17) can be evaluated with the use of (14) and (16) as

$$\begin{aligned} A \sum_{-\nu}^{\nu} j \sinh j\alpha &= A \frac{\partial}{\partial \alpha} \sum_{-\nu}^{\nu} \cosh j\alpha = A \frac{\partial}{\partial \alpha} \frac{\sinh (\nu + \frac{1}{2})\alpha}{\sinh \frac{1}{2}\alpha} \\ &= A \frac{(2\nu + 1) \cosh (\nu + \frac{1}{2})\alpha \sinh \frac{1}{2}\alpha - \sinh (\nu + \frac{1}{2})\alpha \cosh \frac{1}{2}\alpha}{2 \sinh^2 \frac{1}{2}\alpha}, \quad \dots (18) \end{aligned}$$

so that finally, in view of (2)

$$M_n = R \frac{n \sinh \frac{1}{2}\alpha - \cosh \frac{1}{2}\alpha \tanh \frac{1}{2}n\alpha}{4 \sinh^3 \frac{1}{2}\alpha}. \quad \dots (19)$$

Although this formula gives a generally valid solution for the electric polarizability of the chain for any value of n and ω , for several reasons it does not express

the result in a convenient form. Firstly, the parameter z defined in (15) is complex and tables for determining it and the hyperbolic functions of its multiples are not easily available. Furthermore, the final expression would not easily be separable into its real and complex parts for the determination of the components of the polarization in phase and in quadrature with the applied field. Finally, both the numerator and the denominator of (19) tend to zero for vanishing values of ω , and special approximation methods would have to be used to determine the behaviour of the chain at very low frequencies.

A better impression as to the variation of the polarizability-frequency curve with chain length is obtained by working out in detail the formulae for individual cases with small n , where the hyperbolic functions of the multiples of α can be expanded in powers of $\cosh \alpha$ or $\sinh \frac{1}{2}\alpha$. We get

$$(a) \quad n=2, \quad \nu=\frac{1}{2};$$

$$M_2 = \frac{m^2 p_0 F_x}{kT} \frac{1}{2(1 + i\omega\tau_0)} \dots\dots\dots (20)$$

This is equivalent to the case originally suggested by Debye (1929, §20) with $m=2\mu$, $p_0=\frac{1}{2}$; the absorption, determined by the imaginary part of the ratio M_2/F_x , reaches its maximum when $\omega\tau_0=1$.

$$(b) \quad n=3, \quad \nu=1;$$

$$M_3 = \frac{2R}{1 + 2i\omega\tau_0} \dots\dots\dots (21)$$

This is again a Debye curve, but with an effective relaxation time twice that of the previous case and an absorption peak four times as large.

$$(c) \quad n=4, \quad \nu=3/2;$$

$$M_4 = \frac{(5 + 3i\omega\tau_0)R}{1 + 4i\omega\tau_0 + 2(i\omega\tau_0)^2} = R \left(\frac{4.975}{1 + 3.414i\omega\tau_0} + \frac{0.025}{1 + 0.586i\omega\tau_0} \right) \dots\dots (22)$$

$$(d) \quad n=5, \quad \nu=2;$$

$$M_5 = \frac{(10 + 8i\omega\tau_0)R}{1 + 6i\omega\tau_0 + 4(i\omega\tau_0)^2} = R \left(\frac{9.919}{1 + 5.236i\omega\tau_0} + \frac{0.0806}{1 + 0.764i\omega\tau_0} \right) \dots\dots (23)$$

$$(e) \quad n=6, \quad \nu=5/2;$$

$$\left. \begin{aligned} M_6 &= \frac{R}{2} \frac{35 + 56i\omega\tau_0 + 20(i\omega\tau_0)^2}{1 + 9i\omega\tau_0 + 12(i\omega\tau_0)^2 + 4(i\omega\tau_0)^3} \\ &= R \left(\frac{17.32}{1 + 7.464i\omega\tau_0} + \frac{0.167}{1 + i\omega\tau_0} + \frac{0.006}{1 + 0.536i\omega\tau_0} \right) \end{aligned} \right\} \dots\dots (24)$$

The above examples indicate that the polarization-frequency curves consist mainly of a Debye function whose magnitude and effective relaxation time both increase with chain length; superimposed on this are other Debye curves of considerably smaller magnitude which affect only the high frequency tail of the main peak.

A further conclusion can be drawn concerning the behaviour of the chain at very high frequencies for which $\omega\tau_0 \gg 1$. In this case the solution of (12) and (13) is given in first approximation as

$$\varphi_v = 1/2i\omega\tau_0, \quad \varphi_{-v} = -1/2i\omega\tau_0, \quad \varphi_j = 0 \quad (|j| < v), \quad \dots \quad (25)$$

and thus according to (17)

$$M_n = vR/i\omega\tau_0 = (n-1)R/2i\omega\tau_0. \quad \dots \quad (26)$$

This means that at very high frequencies the moment of the chain is proportional to the number of pairs of neighbouring positions. Physically this is explained by the fact that with the rapid variations of the field the probability of transitions between two states is too small to affect appreciably the transitions between either state and its other neighbour. Hence, on an average each chain behaves like a set of $n-1$ dipoles in juxtaposition.

For long chains the solution can be considerably simplified as the range of greatest absorption is shifted to low frequencies for which in view of (15)

$$\omega\tau_0 \ll 1, \quad |\alpha| \ll 1. \quad \dots \quad (27)$$

Then (14) shows that the relative changes of φ_j between neighbouring states are small, the suffix j can be treated as a continuous variable, and the difference equation (12) can be replaced by the differential equation

$$2i\omega\tau_0\varphi(j) = d^2\varphi/dt^2, \quad \dots \quad (28)$$

and the boundary conditions (13) after the transformation indicated in (16) by

$$\frac{d\varphi(v + \frac{1}{2})}{dj} = \frac{d\varphi(\frac{1}{2}n)}{dj} = \frac{d\varphi(-\frac{1}{2}n)}{dj} = 1. \quad \dots \quad (29)$$

The solution of (28) and (29) is given by

$$\varphi(j) = \sinh j\alpha/\alpha \cosh \frac{1}{2}n\alpha, \quad \dots \quad (30)$$

where

$$\alpha = \sqrt{2i\omega\tau_0}. \quad \dots \quad (31)$$

For the calculation of M_n the summation in (17) from $-v$ to $+v$ is replaced by an integration from $-(v + \frac{1}{2}) = -\frac{1}{2}n$ to $+(v + \frac{1}{2}) = +\frac{1}{2}n$ so that

$$\left. \begin{aligned} M_n &= R \int_{-n/2}^{n/2} j\varphi(j) dj = R \int_{-n/2}^{n/2} j \sinh j\alpha / \alpha \cosh \frac{1}{2}n\alpha \\ &= \frac{2R(\frac{1}{2}n \cosh \frac{1}{2}n\alpha / \alpha - \sinh \frac{1}{2}n\alpha / \alpha^2)}{\alpha \cosh \frac{1}{2}n\alpha} = \frac{Rn^3}{4} \frac{\frac{1}{2}n\alpha - \tanh \frac{1}{2}n\alpha}{(\frac{1}{2}n\alpha)^3}. \end{aligned} \right\} \dots \quad (32)$$

This result could have been derived straight from (19) as an approximation valid when (27) holds. The expression (32) is still not suitable for actual computations or a physical interpretation because of the complex argument $\frac{1}{2}n\alpha = \frac{1}{2}n(1+i)\sqrt{\omega\tau_0}$; but it can be given a convenient form by means of the Mittag-Leffler decomposition of $\tanh x$ (cf. Jeffreys and Jeffreys 1946, p. 357). According to this

$$\tanh x = 8x \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2\pi^2 + 4x^2}. \quad \dots \quad (33)$$

With the use of the formula

$$\sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} = \frac{\pi^2}{8}, \quad \dots \quad (34)$$

this yields

$$\frac{\frac{1}{2}n\alpha - \tanh \frac{1}{2}n\alpha}{(\frac{1}{2}n\alpha)^3} = \frac{32}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2 [(2k-1)^2 \pi^2 + n^2 \alpha^2]}, \quad \dots \quad (35)$$

and hence, on substitution for α from (31), (32) becomes

$$M_n = \frac{8Rn^3}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2 [(2k-1)^2 \pi^2 + 2in^2 \omega \tau_0]}, \quad \dots \quad (36)$$

III. RESULTS AND DISCUSSION

The result (36) for the average polarization of a chain for which $n^2 \gg 1$ can be expressed, in view of (6) as

$$M_n = \frac{8m^2 n^3 p_0 E_x}{\pi^4 k T} \left(\frac{1}{1 + i\omega \tau_n} + \frac{1}{3^4} \frac{1}{1 + i\omega \tau_n} + \frac{1}{9} + \frac{1}{5^4} \frac{1}{1 + i\omega \tau_n} + \frac{1}{25} + \dots \right). \quad \dots \quad (37)$$

with

$$\tau_n = 2n^2 \tau_0 / \pi^2, \quad \dots \quad (38)$$

The first term of the series in (37) corresponds to a Debye curve of effective relaxation time τ_n , the remaining terms represent corrections. These give their maximum absolute contribution to the dielectric loss when $\omega \tau_n \approx 9$; at that frequency it amounts to 6 per cent. of the absorption due to the leading term and to 1.3 per cent. of its peak value. The relative contribution due to the correction terms increases with increasing frequency, but never exceeds 25 per cent. of the loss due to the first term.* In the frequency range where the absorption is appreciable ($\omega \tau_n \approx 1$), their contribution is entirely negligible so that the formula (37) is sufficiently well approximated by a single Debye curve

$$M_n = \frac{8m^2 p_0 E_x}{\pi^4 k T} \frac{n^3}{1 + i\omega \tau_n}. \quad \dots \quad (39)$$

For the static case ($\omega = 0$) this expression differs from the exact value (5) by the substitution of $8/\pi^4 = 0.08213$ for $1/12 = 0.08333$ and of n^3 for $n^3 - n$.

The total polarization P per unit volume due to the chain mechanism is obtained by adding the contributions due to the individual chains in the direction of the field; if all the chains are of equal length and oriented in random directions, (39) leads to

$$P = \frac{8m^2 p_0 N_n}{\pi^4 k T} \frac{n^3}{1 + i\omega \tau_n} \langle F \cos^2 \theta \rangle_{\text{av}} = \frac{8N_0 \cdot m^2}{3\pi^4 k T} \frac{n^3}{1 + i\omega \tau_n} \langle F \rangle_{\text{av}}. \quad \dots \quad (40)$$

* For $\omega \tau_n \rightarrow \infty$ the value of the bracket in (37) tends to

$$\frac{1}{i\omega \tau_n} \left(1 + \frac{1}{3^2} + \frac{1}{5^2} + \dots \right) = \frac{\pi^2}{8i\omega \tau_n} = \frac{1.23}{i\omega \tau_n}.$$

Here N_n represents the number of chains and $N_0 = N_n p_0 n$ the number of ions (or dipoles) per unit volume, whereas θ is the angle between the directions of the field and the chain, $\langle \cos^2 \theta \rangle_{av}$ being replaced by $1/3$.

It has so far been impossible to find a definite correlation between an alternating macroscopic field E and the internal field F . In the static case Onsager's (1936) treatment is appropriate for $n=2$, and the theory of Kirkwood (1939) and its extension by Fröhlich (1949, §7) for longer chains; but no generalization to alternating fields exists even of Onsager's theory (an attempt by Fröhlich and Sack (1944) is erroneous). Quantitative conclusions are therefore only possible for substances in which the contribution of the chains to the total polarization is small and no other mechanisms are operative in their absorption range. In this case

$$|\varepsilon - \varepsilon_\infty| \ll \varepsilon_\infty, \quad \dots \quad (41)$$

where ε_∞ is the dielectric constant due to the electronic and atomic polarizations; and if these latter are treated on a macroscopic basis, F and E can be considered equal, so that (40) leads to

$$\varepsilon - \varepsilon_\infty = \frac{32m^2 p_0 N_n}{3\pi^3 kT} \cdot \frac{n^3}{1 + i\omega\tau_n} = \frac{32m^2 N_0}{3\pi^3 kT} \cdot \frac{n^2}{1 + i\omega\tau_n} \quad \dots \quad (42)$$

This formula with (38) shows that both the magnitude of the absorption peak and its effective relaxation time increase with the square of the number of states per chain. If the chains are not of uniform length, then as long as condition (41) applies the contributions of the various chains are additive, so that (40) yields

$$\varepsilon - \varepsilon_\infty = \frac{32m^2}{3\pi^3 kT} \sum \frac{N_n n^3 p_0}{1 + i\omega\tau_n} \quad \dots \quad (43)$$

This result can be described phenomenologically by a distribution of relaxation times τ_n ; in view of the factor n^3 the longer chains will show a greater than proportional influence on the absorption. When the chain density is so high that (41) no longer holds, (42) and (43) are no longer true, but the conclusions drawn from them should still apply in a qualitative way.

The theory needs only slight modification if the system does not contain "chains" but three-dimensional inclusions or "blocks" such that each state is described uniquely not by a scalar dipole moment, but by a vector moment ($j_x m_x, j_y m_y, j_z m_z$) and in an elementary transition any one of the components can change by one unit. In this case the transition probabilities in any direction are unaffected by the field components at right angles and the effects of rectangular blocks on the polarization is a superposition of three chain contributions. If, for instance, the blocks are cubical, and sufficiently far apart for (41) to hold, and if the elementary transition probabilities w_0 and unit components m_x, m_y, m_z are the same in all three directions, the right-hand side of (42) has merely to be multiplied by 3. If the blocks are not rectangular, the effective relaxation time (or times) and the total absorption will still be proportional to the square of their linear dimension as in (42), but will also depend on their shape.

Considerations similar to those preceding and following (43) apply when the blocks are of varying sizes or shapes or sufficiently concentrated for condition (41) to break down.

It is of interest to note that if the polarization is due to moving charges the differential equation arising as an approximation of the original difference equation (9) can be written as

$$\partial \rho / \partial t = w_0 a^2 \partial^2 \rho / \partial s^2 = D \partial^2 \rho / \partial s^2, \quad \dots \dots \dots (44)$$

where a denotes the distance between two equilibrium positions of ions corresponding to two successive "states", $s = s_0 + ja$ is the linear coordinate measured along the chain and $\rho = p/a^3$ the probability density for the charged carriers. Equation (44) can arise from a number of other mechanisms, more general than the special model expressed by (9), provided the diffusion coefficient D is suitably interpreted. In this sense the theory presented here is similar to one developed by Jaffé (1933) who considers the effect of ions which can move across a solid under the action of a given potential difference. In Jaffé's theory the range of movement of the ions is limited by the electrodes only; he further takes into account the continuous generation and recombination of ions and is mainly interested in saturation effects, which are completely neglected in this paper; on the other hand he does not consider the behaviour of such a system under the action of an alternating field.

IV. APPLICATIONS

The main conclusion to be drawn from this theory is that whenever the dielectric polarization and absorption can be expressed in terms of a chain mechanism, the experimentally observed relaxation time, or its mean value, does not give a true measure of the probability of an elementary transition; according to (38) it can exceed the inverse of the transition probability by several orders of magnitude. The activation energy H for an elementary process is usually determined from the variation of the mean relaxation time τ with temperature according to the formula

$$\tau = (A\pi/2\omega_a) \exp (H/kT), \quad \dots \dots \dots (45)$$

where ω_a is the oscillation frequency of an elementary unit (ion or dipole) and A a factor connected with the entropy of activation (cf. Fröhlich 1949, equation 11.3). In our case the factor $2n^2 \mp 2$ has to be taken into account. Its disregard, if the temperature variation of n is sufficiently rapid, would lead to erroneous values of H and in any case it would seriously affect the pre-exponential factor A .

The "block" model described in Section III is applicable to the dielectric properties of ionic crystals containing lattice imperfections which can move over a certain distance in the crystal. These considerations apply whether the imperfections are vacant lattice sites or interstitial ions, and the elementary dipole in either case is given by $ea \varepsilon_\infty$, where e represents the charge on the ion (missing or interstitial) and a the distance moved in an elementary step (in most cases the lattice parameter). If the number of cells over which the defect can move is reasonably uniform all through the material, a single macroscopic relaxation time is associated with each type of imperfection. The absorption

peaks observed at comparatively low frequencies by Breckenridge (1948, 1950) could be explained by such a mechanism; a detailed comparison between theory and experiment is difficult as in the few cases where the absorption has been measured as a function of frequency at constant temperature an insufficient number of points are shown to determine the exact shape of the curve, and for variable temperature the present theory can be applied only if the temperature dependence of the average range of motion is known. If the explanation for the losses is correct, for the reasons stated above the determination of activation energies and entropies from experimental results must take into account corrections arising from the chain mechanism.

As mentioned in the Introduction it has been suggested in an earlier paper (Meakins and Sack 1951) that the high dielectric constants and loss factors of certain forms of solid long-chain alcohols are to be explained by means of a linear-chain mechanism. Owing to the crystal structure of these substances their hydroxyl groups lie in definite layers, and their directions will be stabilized by hydrogen bonds to neighbouring OH-groups if their distances apart are between 2.5 and 2.8 Å. Successive hydrogen bonds form a chain, the length of which is determined by misfits or crystal imperfections. For steric reasons, arrangements in which two hydrogen atoms face each other such as



are impossible, so that within each chain there can be at most one pair of anti-parallel neighbours of the type



This discontinuity can wander along the chain through reversal of the adjoining hydroxyl groups, a displacement over one molecule producing a change of approximately $2\mu_{\text{OH}}/\epsilon_\infty$ in the total moment of the chain where μ_{OH} is the dipole moment of an OH-group. If all the dipoles in a chain are parallel, this corresponds to the discontinuity passing right out of it; from energy considerations these two states are more favoured than the remainder as they contain one additional hydrogen bond; but this merely somewhat modifies the boundary conditions (13) and hence (29) without affecting the general nature of the results.

An alternative mechanism suggested to account for the dielectric properties of substances containing hydrogen bonds is that each proton can take up two equilibrium positions nearer to one or other of its adjoining oxygen atoms and can pass between these across the intervening energy barrier, but that no two hydrogen atoms are attached to the same oxygen at the same time. This rules out arrangements of the type (C)



and only allows for one discontinuity of the type (D)



in each chain. This mechanism requires that one of the protons is separated from the chain. In this case the movement of the discontinuity by one link alters the dipole moment of the chain approximately by the product of the effective charge and the displacement of the hydrogen atom.

The wide frequency range of the absorption in solid long-chain alcohols indicates that the chains vary considerably in length. Which of the two mechanisms is mainly operative has not been determined. If the ends of the chains, instead of presenting an impenetrable barrier (as expressed in equation (10)) allow ions to pass on to further positions with a greatly reduced transition probability, then both mechanisms can coexist leading to a D.C. conductivity by means of proton transfer; such a combined mechanism has been suggested by Stearn and Eyring (1937) to account for the conductivity of water: confirmation that the conduction in solid primary alcohols is due to hydrogen ions has been obtained by Kakiuchi, Komatsu, and Kyoya (1951).

The high D.C. conductivity and dielectric dispersion at low frequencies observed in the α -forms of monoglycerides of fatty acids by Crowe and Smyth (1950) can be accounted for by the same bond chain mechanism, though these authors explain the losses in terms of molecular rotations.

V. LIMITATIONS OF THE THEORY

It should be emphasized that the theory developed here only accounts for the properties of chains if they are straight and each value of their total effective dipole moment is associated with a single state only. If antiparallel neighbours of the type (A) or (C) are possible, the model proposed by Ising (1925) applies for which the induced moment per chain is proportional to its length and the effective relaxation time independent of it (Sack 1946). If the probability of dipoles facing each other is small but finite, the behaviour of moderately short chains will be described by the theory of this paper, but with increasing chain length the conclusions based on Ising's model will be valid. Similarly, if the chain lengths are not straight, but kinked, their effective polarizability per dipole (Kirkwood 1939) and presumably their relaxation time are independent of chain length. This explains the increase in the static dielectric constant when certain substances containing hydrogen bonds are solidified from the liquid state (see, for example, Hoffman and Smyth 1949): in the solid the crystal structure forces the bond chains to develop in straight lines, whereas in the liquid they are randomly kinked.

If the chains are branched, again the theory does not apply quantitatively, but the general conclusion can be inferred that, as in linear chains, the absorption is greatly enhanced and shifted to low frequencies. This would account for the very high static dielectric constant of ice and its absorption peak at comparatively low frequencies (see, for example, Lamb 1946), and for similar properties in amorphous sugars (Alexandrov, Kobeko, and Kuvshinski 1936) and supercooled glycerine (Morgan 1934): in these substances the hydrogen bonds are known or can be assumed to form three-dimensional networks. The spontaneous polarization and high dielectric constant of potassium dihydrogen phosphate has been explained in a similar manner in terms of a proton transfer in a network of hydrogen bonds between the PO_4 groups (Slater 1941).

VI. REFERENCES

- ALEXANDROV, B., KOBOKO, P., and KUVSHINSKI, E. (1936).—*Tech. Phys. U.S.S.R.* **3** : 495.
- BAKER, W. O., and SMYTH, C. P. (1938).—*J. Amer. Chem. Soc.* **60** : 1229.
- BRECKENRIDGE, R. G. (1948).—*J. Chem. Phys.* **16** : 959.
- BRECKENRIDGE, R. G. (1950).—*J. Chem. Phys.* **18** : 913.
- CROWE, R. W., and SMYTH, C. P. (1950).—*J. Amer. Chem. Soc.* **72** : 4427.
- DEBYE, P. (1929).—"Polar Molecules." (The Chemical Catalog Co. : New York.)
- FRÖHLICH, H. (1949).—"Theory of Dielectrics : Dielectric Constant and Dielectric Loss." (Clarendon Press : Oxford.)
- FRÖHLICH, H., and SACK, R. (1944).—*Proc. Roy. Soc. A* **182** : 388.
- HOFFMAN, J. D., and SMYTH, C. P. (1949).—*J. Amer. Chem. Soc.* **71** : 431.
- ISING, E. (1925).—*Z. Phys.* **31** : 253.
- JAFFÉ, G. (1933).—*Ann. Phys. Lpz.* **16** : 217.
- JEFFREYS, H., and JEFFREYS, B. S. (1946).—"Methods of Mathematical Physics." (Cambridge Univ. Press.)
- KAKIUCHI, Y., KOMATSU, H., and KYOYA, S. (1951).—*J. Chem. Phys.* **19** : 132.
- KIRKWOOD, J. G. (1939).—*J. Chem. Phys.* **7** : 911.
- LAMB, J. (1946).—*Trans. Faraday Soc.* **42A** : 238.
- MEAKINS, R. J., and MULLEY, JOAN W. (1951).—*Aust. J. Sci. Res. A* **4** : 365.
- MEAKINS, R. J., and SACK, R. A. (1951).—*Aust. J. Sci. Res. A* **4** : 213.
- MEAKINS, R. J., and WELSH, H. K. (1951).—*Aust. J. Sci. Res. A* **4** : 359.
- MORGAN, S. O. (1934).—*Trans. Electrochem. Soc.* **65** : 109.
- ONSAGER, L. (1936).—*J. Amer. Chem. Soc.* **58** : 1486.
- SACK, R. A. (1946).—*Trans. Faraday Soc.* **42A** : 61.
- SLATER, J. C. (1941).—*J. Chem. Phys.* **9** : 16.
- STEARNS, A. E., and EYRING, H. (1937).—*J. Chem. Phys.* **5** : 113.

THE VISCOSITY AND RIGIDITY OF GELATIN IN CONCENTRATED AQUEOUS SYSTEMS

I. VISCOSITY

By C. W. N. CUMPER* and A. E. ALEXANDER*

[*Manuscript received August 29, 1951*]

Summary

The viscosity of gelatin solutions at concentrations between 100 and 500 g./l. has been studied as a function of pH, ionic strength, and temperature. In these concentrated systems the predominant factor determining the viscosity appeared to be Coulombic forces between ionized groups on adjacent chains.

I. INTRODUCTION

Arising out of the authors' earlier work on the surface denaturation of proteins, it became essential to have certain information about concentrated solutions and gels, particularly for proteins of very asymmetric shape. Gelatin fulfils this asymmetry requirement very well for its molecule is about 18 Å in width and approximately 300 to 800 Å in length, depending upon the molecular weight (e.g. see Ferry 1948, p. 1). The effect of pH, temperature, and ionic strength on the viscosity and rigidity of gelatin over the concentration range from 100 to 500 g./l. was therefore studied. The results are discussed here but their bearing on adsorbed protein films will be deferred until a subsequent paper.

One characteristic of the viscosity of dilute gelatin solutions seems to be that it either increases or decreases with time. Above the gelation temperature the viscosity usually decreases slowly at pH values other than the isoelectric point. No time effects were detected over periods of about 30 minutes except at the very highest concentrations used here (>400 g./l.), when the viscosity increased slightly over a period of 15 minutes. With solutions more viscous than about 30 poises the viscosity decreased slightly with increasing rate of shear; with other solutions the flow appeared to be Newtonian. Structural rigidity appeared to be absent at the temperature usually employed (40°C.).

II. EXPERIMENTAL

The sample of gelatin was prepared by Davis Gelatine (Aust.) Pty. Ltd., and had an ash content of 0.05 per cent. and an isoelectric point of 5.0. To characterize it, the intrinsic viscosity $[\eta]$ was determined using a capillary flow viscometer at 35°C. (above the gelation temperature), with 1 per cent. of salt (NaCl) added to eliminate electroviscous effects. The results, shown in Figure 1, gave the intrinsic viscosity as 0.31 which, using the results of Scatchard

* School of Applied Chemistry, N.S.W. University of Technology, Sydney.

et al. (1941), corresponds to a molecular weight of 31,000. This intrinsic viscosity also corresponds to an axial ratio of 21, from the equation of Simha (1940), for an ellipsoid of revolution, and hence the dimensions of the molecule in a dilute solution are about $18 \times 380 \text{ \AA}$.

The amino-acid composition of the gelatin was not determined but, according to the published literature, gelatin is characterized by a high content of glycine, proline, and hydroxyproline (Cohn and Edsall 1943, p. 368). These account

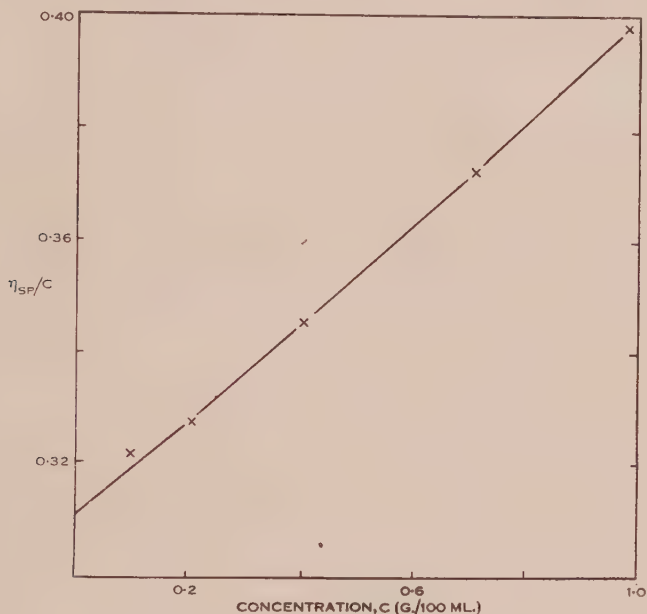


Fig. 1.—The intrinsic viscosity of gelatin as a function of the concentration.

for 60 per cent. of the total number of amino-acid residues; 7 per cent. of the residues contain acidic groups in their side-chains and 11 per cent. contain basic groups.

The viscosities of the concentrated solutions were determined using a Couette type viscometer, details of which have been published elsewhere (Gray and Alexander 1949). The gap between the cylinders was 2 mm. and the radius of the inner one 2.5 cm. The angular velocity of the outer cylinder was varied between 0.80 and 4.4 radn./sec., and the inner cylinder (length 8.0 cm.) was always completely immersed to avoid errors arising from "skin" formation. Under the conditions used the correction for "end effects" was negligible.

If the gelatin was dissolved by allowing it to swell slowly in water, it was impossible to prevent frothing; it was therefore dissolved by heating to 50 to 55 °C. for 15 minutes which should cause little degradation. Dissolving at higher temperatures caused a definite lowering of the viscosity.

Unless otherwise stated the isoelectric gelatin was dissolved in distilled water and the viscosities measured at 40 ± 0.1 °C.

III. RESULTS

(a) *Effect of Concentration*

The viscosity over the concentration range from 150 to 480 g./l. is shown in Figure 2. It increased very rapidly at the higher concentrations but the plot of $\log \eta_{sp}/C$ against C gave a graph of small curvature which can be represented quite accurately by the empirical equation

$$\log \eta_{sp}/C = -1.51 + 4.3 \times 10^{-3}C + 3.5 \times 10^{-6}C^2.$$

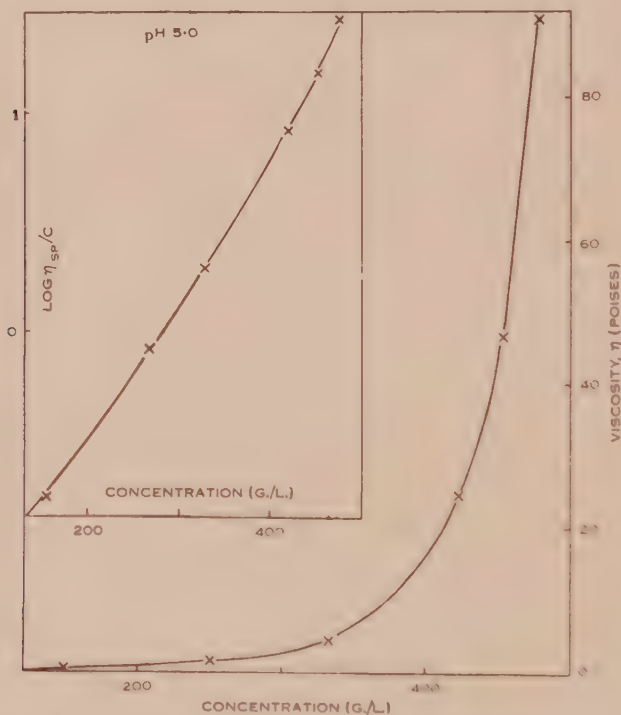


Fig. 2.—The viscosity of gelatin as a function of concentration.

The $\lim_{C \rightarrow 0} \eta_{sp}/C$ is 0.032 (C in g./l.) in excellent agreement with the value determined directly in dilute solutions, indicating that there is no fundamental change in the nature of the gelatin solution over the whole concentration range.

(b) *Effect of pH*

The pH (measured by a glass electrode) of a solution containing 270 g./l. was adjusted to suitable values by adding strong acid or alkali. The viscosity-pH curve of the resulting system (Fig. 3) shows a rather broad maximum around the isoelectric point, the viscosity falling away sharply on either side so that at pH 1 and 10 it was little greater than that of water. pH changes over the range from 3.5 to 8 were reversible but outside this range the viscosity was always lower on returning to the isoelectric point. After returning from pH 1, for instance, the viscosity was only 10 per cent. of its correct value even after allowing for the

increased salt content and standing for 24 hours. Some permanent change, probably degradation, had evidently occurred.

In contrast to the above results the viscosity of dilute gelatin solutions (*c.* 8 per cent.) passes through a rather sharp minimum at the isoelectric point and has a maximum on either side (*e.g.* see Sheppard and Houck 1930).

(c) *Effect of Ionic Strength*

At high concentrations the viscosity of isoelectric gelatin was found to be reduced by the addition of sodium chloride up to an ionic strength of 1.5 (Fig. 4). With higher salt concentrations the viscosity increased, probably due to the "salting out" of the gelatin and the consequent tendency to aggregation. In fact it was found difficult to dissolve the gelatin in strong salt solutions as it tended to form coacervates.

(d) *Effect of Temperature*

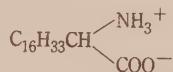
With increasing temperature the viscosity of the concentrated solutions fell rapidly up to about 43 °C., after which it decreased more slowly. This is shown in Figure 5, where the logarithm of the viscosity is plotted against the reciprocal of the absolute temperature. Above 43 °C. the solutions flow less freely than anticipated from the behaviour at lower temperatures.

IV. DISCUSSION

In concentrated gelatin solutions the resistance to flow will arise largely from intermolecular interactions of the following origin :

- (i) Coulombic forces between oppositely charged ions (presumably mainly -COO^- and -NH_3^+).
- (ii) Ion-dipole forces (*e.g.* -COO^- on one chain and -NH_2 on the other ; or -NH_3^+ and -COOH).
- (iii) Hydrogen bonding between >C=O and H-N< .
- (iv) London forces, particularly between the non-polar side-chains.
- (v) Mechanical entanglements between the long chains as found with non-polar polymers in ideal solvents.

The experimental findings that added salts greatly reduce the viscosity (Fig. 4) and that the viscosity is a maximum at the isoelectric point (Fig. 3) can only readily be explained if the Coulombic term far outweighs the others in importance at these high concentrations. This is not unexpected on general grounds and is in keeping with the conclusions drawn from the study of appropriate monolayer systems, *e.g.* monolayers of



(Alexander 1942). Assuming that the viscosity is determined by the -COO^- , H_3^+N -interactions on neighbouring chains it should be possible to construct the η -pH curve from the titration curve of gelatin. The curve thus

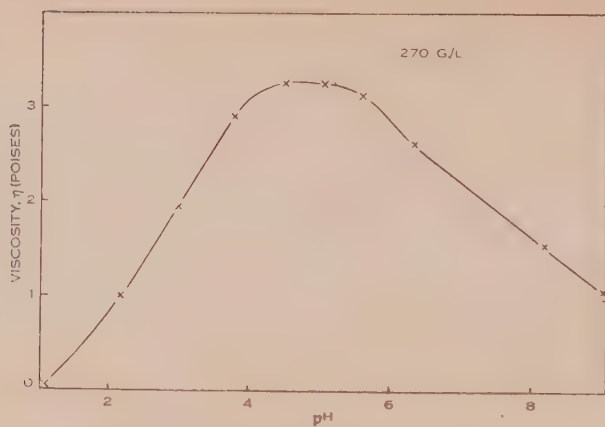


Fig. 3.—The viscosity of gelatin (270 g./l.) as a function of pH.

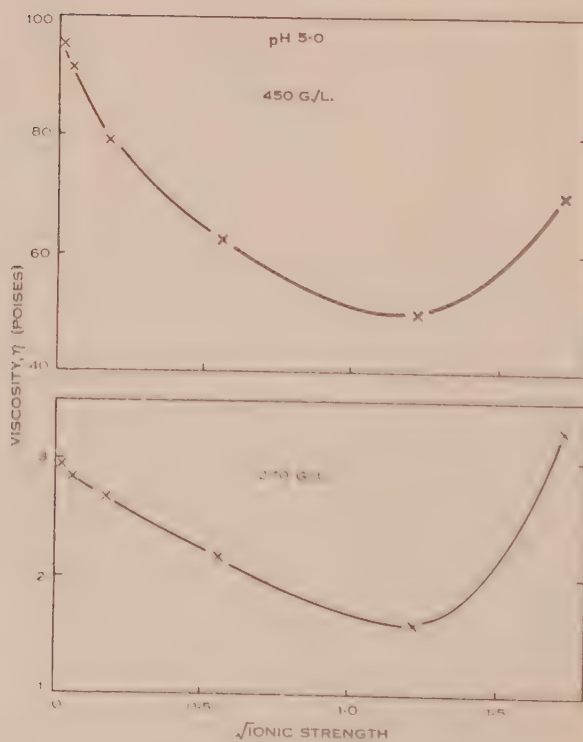


Fig. 4.—The viscosity of gelatin at two concentrations as a function of the square root of the ionic strength.

calculated shows the general shape of the experimental curve but the agreement is not quantitative, probably due to the varying accessibility and strength of the charged groups concerned.

The reduction in viscosity produced by added salt (see Fig. 4) also finds a ready explanation in terms of Coulombic forces, for according to simple Debye-Hückel theory we have

$$F = F_0(1 - \kappa r),$$

where F_0 = force in absence of added salt,

F = force in presence of added salt of ionic strength I ,

κ = Debye-Hückel function ($\kappa = 0.327 \times 10^8 \sqrt{I}$),

r = average separation of the charged centres.

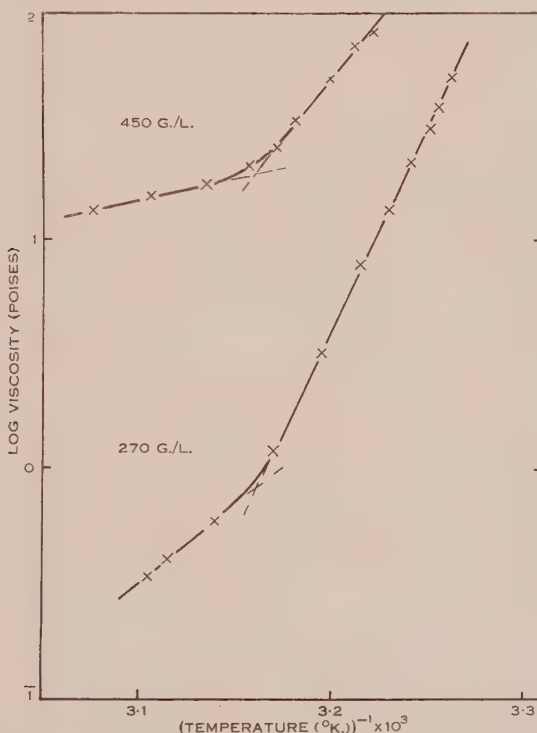


Fig. 5.—The viscosity of gelatin at two concentrations as a function of the reciprocal of the absolute temperature.

At the lower gelatin concentrations η is approximately proportional to \sqrt{I} over a considerable range, as predicted by the above equation, and from this we calculate r to be about 1.5 Å. This is not an unreasonable value although somewhat smaller than expected. At the higher gelatin concentration the η - \sqrt{I} relationship is less linear but this is not surprising in view of the approximate nature of the above equation. The upward swing at high salt concentrations is undoubtedly due to a salting out phenomenon, as mentioned earlier.

The rather unexpected temperature variation of the viscosity, particularly the break at *c.* 43 °C. (see Fig. 5), remains to be considered. Gelation effects are unlikely since the gelation temperature of the present sample was *c.* 31 °C. The effect of temperature was reversible and the transition temperature independent of the gelatin concentration, thus precluding any explanation based upon changes in intermolecular packing. An intramolecular change, such as an increased flexibility of the chains leading either to increased mechanical entanglements, or a closer average approach of the oppositely charged ionized groups, would thus seem to be indicated.

V. ACKNOWLEDGMENTS

We would like to thank Davis Gelatine (Aust.) Pty. Ltd. for a gift of the gelatin used, and one of us (C.W.N.C.) would also like to thank the Goldsmiths' Company for the award of a travelling scholarship.

VI. REFERENCES

- ALEXANDER, A. E. (1942).—*Proc. Roy. Soc. A* **179**: 470.
COHN, E. J., and EDSALL, J. T. (1943).—"Proteins, Amino Acids and Peptides." (Rheinhold Publ. Corp.: New York.)
FERRY, J. D. (1948).—"Advances in Protein Chemistry." Vol. 4. (Academic Press: New York.)
GRAY, V. R., and ALEXANDER, A. E. (1949).—*J. Phys. Colloid. Chem.* **53**: 9.
SCATCHARD, G., ONCLEY, J. L., WILLIAMS, J. W., and BROWN, A. (1944).—*J. Amer. Chem. Soc.* **66**: 1980.
SHEPPARD, S. E., and HOUCK, R. C. (1930).—*J. Phys. Chem.* **34**: 273.
SIMHA, R. (1940).—*J. Phys. Chem.* **44**: 25.

THE VISCOSITY AND RIGIDITY OF GELATIN IN CONCENTRATED AQUEOUS SYSTEMS

II. RIGIDITY

By C. W. N. CUMPER* and A. E. ALEXANDER*

[Manuscript received August 29, 1951]

Summary

The rigidity of the gels formed by cooling the systems considered in Part I of this series was also measured. The results obtained have been explained by postulating a network of the gelatin molecules held together by electrostatic bonds.

I. INTRODUCTION

The rigidity of gelatin gels has been studied extensively (Ferry 1948*a*, p. 1) but here again little has been done at high concentrations. Gelation has been attributed to a linkage of the elongated molecules into a three-dimensional network by intermolecular forces. No theoretical equation for the rigidity has yet been formulated, but Ferry and Eldridge (1949) found that the empirical equation

$$n^{1/2}/c = 1.22 \times 10^{-4} (M_w - 3.1 \times 10^{10} e^{-7900/RT})$$

described how the modulus of rigidity (n) of their gelatins depended upon concentration (c), weight average molecular weight (M_w), and temperature (T).

In contrast to the viscosity determinations the rigidity depended markedly upon the time factor increasing rapidly at first and then slowly over a very long period. The results reported here are for the rigidity taken shortly after the end of the first stage. When the rigidity of the gel at 25 °C. was 10^4 dynes/cm.² for example, the solution, on cooling from 50 °C., required about 24 hours to attain this value; if it was 10^6 dynes/cm.², the rigidity was essentially constant within 3 hours. A longer time was required at lower temperatures.

II. EXPERIMENTAL

The shear rigidity of the gels was measured by the concentric cylinders method (Alexander and Johnson 1949, p. 474). The radii of the two cylinders were 0.45 and 0.75 cm. and the depth of immersion about 9 cm. The rigidity modulus did not depend upon the depth of immersion.

The same gelatin was used as before and, unless otherwise stated, the rigidities were measured at 25 ± 0.05 °C. and at the isoelectric point (pH 5.0).

* School of Applied Chemistry, N.S.W. University of Technology, Sydney.

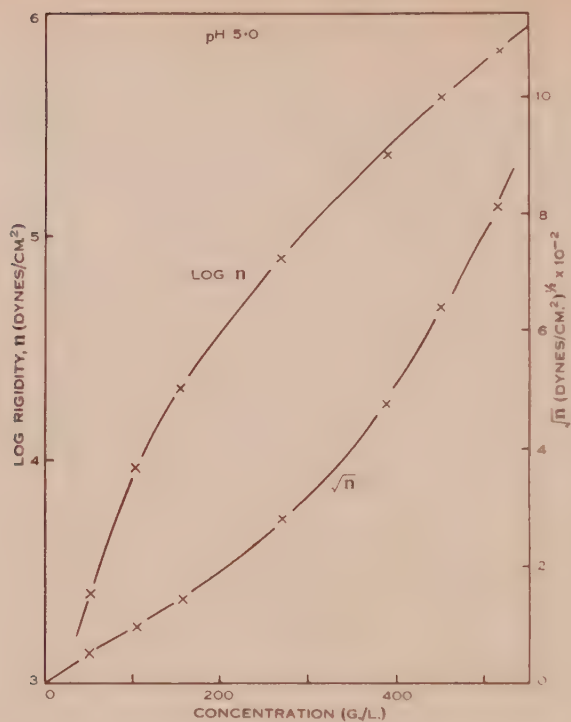


Fig. 1.—The rigidity of gelatin as a function of concentration.

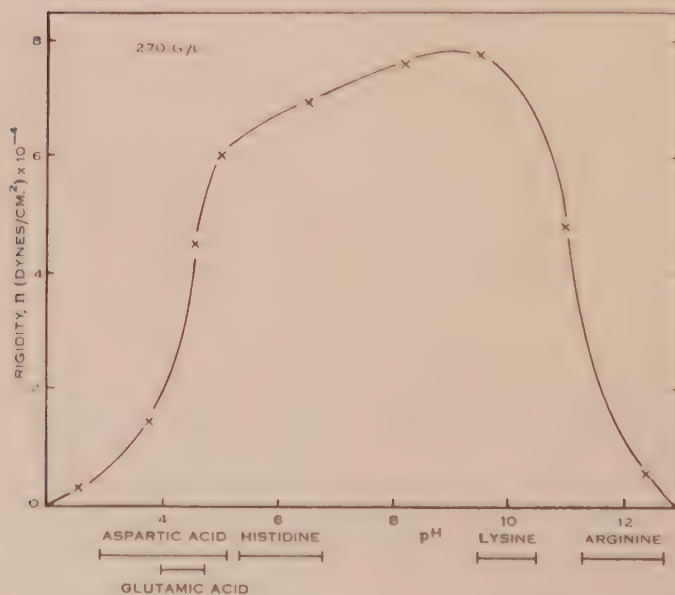


Fig. 2.—The rigidity of gelatin (270 g./l.) as a function of pH.

III. RESULTS

(a) *Effect of Concentration*

The rigidities of gels with concentrations between 50 and 520 g./l. are shown graphically in Figure 1. Ferry's equation predicts, and as found in many investigations, that the plot of \sqrt{n} against c should be a straight line. This was found to be so up to a concentration of about 200 g./l. but above this point the rigidity increased more rapidly.

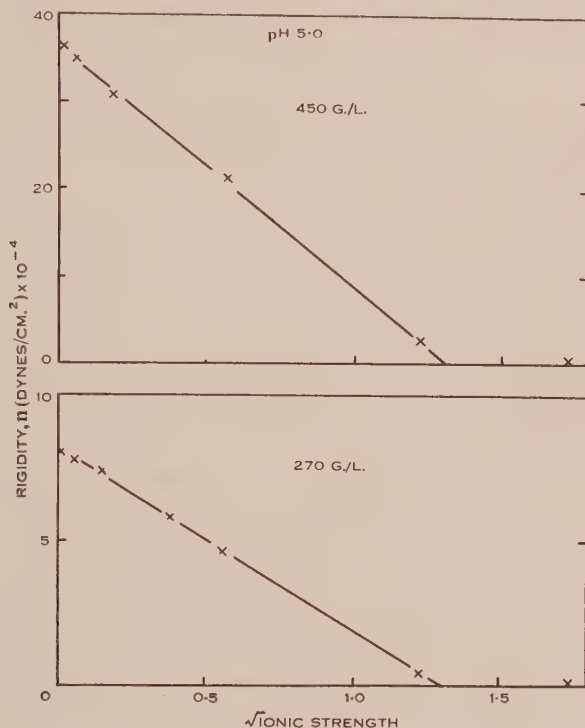


Fig. 3.—The rigidity of gelatin at two concentrations as a function of the square root of the ionic strength.

The application of Ferry's equation at the lower concentrations gives a value of 61,900 for the weight average molecular weight (M_w). This is considerably larger than the value obtained from the intrinsic viscosity (31,000) but the same order of divergence has been found for gelatin by other workers (for example, Ferry 1948*b*).

(b) *Effect of pH*

The rigidity of a gel containing 270 g./l. gelatin at a constant ionic strength of 0.15 shows an unexpected dependence on pH (Fig. 2), with a maximum at pH 9 and *not* at the isoelectric point. However, a change in slope can be seen close to the isoelectric point. At low concentrations Gerngross (1926) found that the rigidity was practically independent of pH over a considerable range near the isoelectric point.

(c) *Effect of Ionic Strength*

At low gelatin concentrations Northrop and Kunitz (1926) and Bungenberg de Jong and Henneman (1932) found that, in general, the rigidity was decreased by the presence of salt. Figure 3 shows that at both concentrations studied here, the rigidity varies linearly with the square root of the ionic strength, the relation being

$$n = n_0(1 - 0.77\sqrt{I}),$$

where n_0 is the rigidity in absence of added salt.

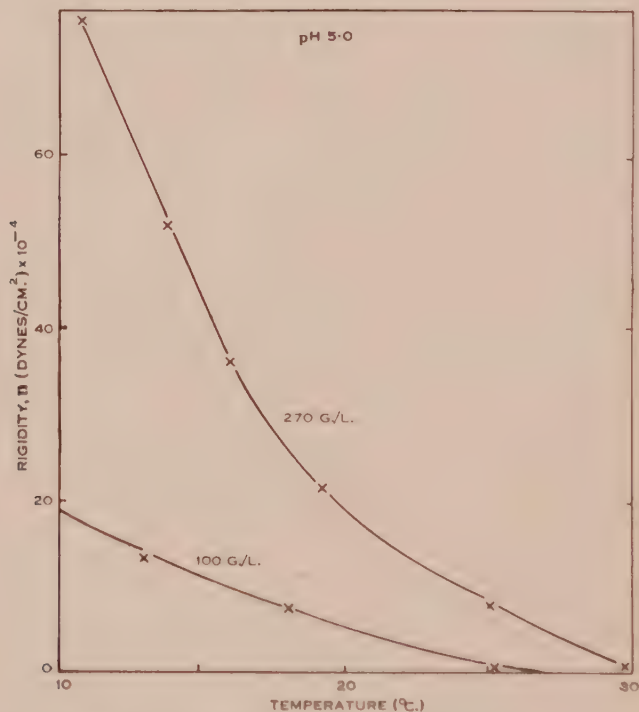


Fig. 4.—The rigidity of gelatin at two concentrations as a function of the temperature.

(d) *Effect of Temperature*

The results for the effect of temperature at gelatin concentrations of 100 and 270 g./l. are shown in Figure 4. The curve for the lower concentration fits Ferry's equation very well but at a concentration of 270 g./l. the rigidity is much greater and cannot be made to fit the equation by changing just one of its constants.

IV. DISCUSSION

It would be expected that essentially the same factors would determine the rigidity and viscosity behaviour. From the viscosity data of Part I of this series, it seemed that the Coulombic forces between oppositely charged ionic groups on neighbouring chains were by far the most important factors and this

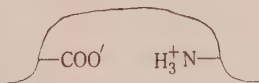
is borne out by the rigidity data. Thus in a general way the variations of rigidity with concentration, pH, ionic strength, and temperature, parallel the variations in viscosity.

The evidence which Ferry (1948*a*, p. 1) has compiled postulating a network structure in the gel seems overwhelming. From the properties of dilute gelatin gels (<10 per cent.) he favours a non-ionic associating force; probably strong van der Waals attraction between several amino-acid residues at each linkage. His main evidence was the small influence of pH and salt upon the rigidity and melting point of the gel, but we feel that the experimental results do not exclude an explanation based upon electrostatic forces.

The fundamental difference between the influence of Coulombic forces upon the viscosity and rigidity of gelatin is as follows: *All* the ionized side-chains will contribute to the viscosity (the system being mobile, over a sufficient time all possible types of side-chain contact will occur, and the observed effect will thus be an average one) whereas the rigidity will be determined *solely* by those *few* side-chains that actually form intermolecular salt bridges; those formed being the ones of lowest energy consistent with steric restrictions.

The total number of ionic bonds formed in the gel will, for several reasons, exceed those contributing to the rigidity:

- (i) Some will be on molecules which have formed only one intermolecular linkage, e.g. $R \cdot COO'Na^+$ or $R \cdot NH_3^+Cl^-$.
- (ii) Salt linkages are non-directional so the existence of a bond does not necessarily imply that it will contribute to the shear rigidity. Only those opposing the strain in the particular direction in which it is applied are effective.
- (iii) Some bonds will be intra- rather than intermolecular, e.g.



(a) Effect of Ionic Strength

Let us first consider the variation of rigidity with ionic strength which, as pointed out earlier, obeys the relationship

$$n = n_0(1 - 0.77\sqrt{I}).$$

From the Debye-Hückel treatment (Cumper and Alexander 1952, p. 151), assuming $F \propto n$, we obtain

$$n = n_0(1 - \alpha r),$$

where the symbols have the same meaning as before. Hence r (the average separation of the charged centres) is found to be 2.4 Å. This value would seem to be small but not unreasonably so in view of the Debye-Hückel approximations. The viscosity data gave a value of the same order,

(b) *Effect of pH*

A gelatin molecule of molecular weight 40,000 would, from its chemical analysis (Cumper and Alexander 1952, p. 147), contain approximately 30 acidic and 47 basic amino-acid residues. Not all these are capable of forming ionic bonds, however, firstly, because of steric factors, and secondly, because of the necessity for ionization at the pH in question.

Steric factors could arise in two ways :

- (i) From the salt linkages which will render useless any ionic groups within a certain distance, due to the finite width of the gelatin molecule.

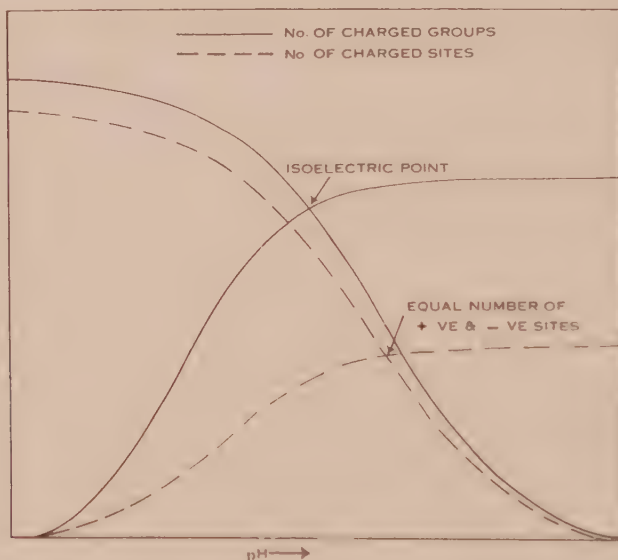


Fig. 5.—Variation in number of positively and negatively charged groups and sites in gelatin with pH (diagrammatic).

- (ii) From a difference in the distribution of the acidic and basic amino-acid residues along the polypeptide chain. There are fewer acidic than basic amino-acid residues in gelatin and if we assume, as found for wool (Consden 1949 ; Consden, Gordon, and Martin 1949) and insulin (Sanger 1949), that there is a strong tendency for the acidic amino-acid residues to occur in pairs while the basic ones occur singly, then the number of positively and negatively charged *sites* depend upon pH as depicted by the broken lines in Figure 5. A pH above the isoelectric point, and not the isoelectric point itself, would then be the “ unique ” point of the rigidity-pH curve.

As regards the ionization factor the only residues to be considered are aspartic, glutamic, and hydroxyglutamic acids, histidine, lysine, and arginine. The range of pK values exhibited by these residues in proteins is shown below the rigidity-pH curve in Figure 2 (Cohn and Edsall 1943, p. 445).

As already stressed, few of the charged sites would actually be involved in intermolecular salt bridges. At high pH's when the pK values of lysine and, more particularly, of arginine are exceeded, the number of basic amino-acid residues available for bond formation decreases very rapidly and consequently there is a sharp fall in rigidity between pH's 10 and 12. Similarly the ionization of aspartic acid would control the decrease in rigidity between pH's 4.5 and 2.

The experimental data also indicate that between these two extremes there is no major change in the amino acids involved in bond formation. The gradual increase in rigidity between pH's 5 and 9 would indicate either an actual increase in the number of bonds or the progressive replacement of some bonds by others of greater stability.

(c) Effect of Temperature

An increase of temperature would increase the vibration of the chains and hence decrease the rigidity but there does not seem to be any suitable mathematical equation which fits the experimental results.

(d) Effect of Concentration

At gelatin concentrations below about 200 g./l. the rigidity is proportional to the square of the concentration as one would expect (Ferry 1948*b*). At higher concentrations the number of effective intermolecular linkages apparently increases more rapidly than this square law predicts, or some other factor such as the increased mechanical entanglements of the chains becomes of importance.

V. ACKNOWLEDGMENTS

We would like to thank Davis Gelatine (Aust.) Pty. Ltd. for a gift of the gelatin used, and one of us (O.W.N.O.) would also like to thank the Goldsmiths' Company for the award of a travelling scholarship.

VI. REFERENCES

- ALEXANDER, A. E., and JOHNSON, P. (1949).—“Colloid Science.” (Oxford Univ. Press.)
BUNGENBERG DE JONG, H. G., and HENNEMAN, J. P. (1932).—*Kolloidzshr.* **36**: 123.
COHN, E. J., and EDSALL, J. T. (1943).—“Proteins, Amino Acids and Peptides.” (Reinhold Publ. Corp.: New York.)
CONSDEN, R. (1949).—*J. Text. Inst. Manchr.* **40**: 814.
CONSDEN, R., GORDON, A. H., and MARTIN, A. J. P. (1949).—*Biochem. J.* **44**: 584.
CUMPER, C. W. N., and ALEXANDER, A. E. (1952).—*Aust. J. Sci. Res. A* **5**: 146.
FERRY, J. D. (1948*a*).—“Advances in Protein Chemistry.” Vol. 4. (Academic Press: New York.)
FERRY, J. D. (1948*b*).—*J. Amer. Chem. Soc.* **70**: 2244.
FERRY, J. D., and ELDRIDGE, J. E. (1949).—*J. Phys. Colloid. Chem.* **53**: 184.
GERNGROSS, O. (1926).—*Kolloidzshr.* **40**: 279.
NORTHROP, J. H., and KUNITZ, M. (1926).—*J. Gen. Physiol.* **8**: 317.
SANGER, F. (1949).—*Biochem. J.* **45**: 363.

OLEIC ACID MONOLAYERS ON CONCENTRATED SALT SOLUTIONS

By A. R. GILBY* and THE LATE E. HEYMANN*

[*Manuscript received August 20, 1951*]

Summary

Equilibrium spreading pressures (E.S.P.) and force-area (π - A) curves have been measured for films of oleic acid spread on solutions of a range of salts and non-ionic compounds at various concentrations. There is a linear relationship between E.S.P. and activity of salt in the subphase up to an activity of about 5. The films are still of the liquid expanded type on salt solutions but the π - A curves are shifted to larger areas. Ions follow the order of the lyotropic series in the magnitude of their effects, anions having a greater effect than cations. Results are explained by visualizing both a hydration effect and a charge effect causing an increase in attraction between oleic acid and the substrate, together with an increase in the repulsive terms of the system of forces between the film molecules.

I. INTRODUCTION

That polyvalent ions in an alkaline subphase often change the mechanical properties of a fatty acid monolayer was stressed by Myers and Harkins (1937). Mitchell, Rideal, and Schulman (1937) have pointed out that such ions affect the chemical as well as the physical properties of the film as shown, for example, by study of reaction rates in surfaces. By analysing skims of stearic acid on dilute solutions of calcium and barium salts at varying pH, Langmuir and Schaefer (1936, 1937) showed that these phenomena have a chemical basis due to the formation of varying proportions of insoluble soap. Salt concentrations were of the order of 10^{-4} molar.

However, working on more concentrated salt subphases, Harkins and Morgan (1925) showed that, when chemical effects are absent, the presence of a salt causes an increase in the area of the monolayer. They found further that certain very slightly polar substances spread on concentrated calcium chloride solutions whereas they did not spread on water. Thus the polar character of the subphase was apparently accentuated and the attraction between film-forming substance and the subphase increased. Adam, Askew, and Pankhurst (1939) investigated spreading on a 34 per cent. ammonium sulphate solution. The surface pressure at constant area per molecule was increased on the salt solutions but the shape of the force-area (π - A) curves was unchanged. In a more comprehensive investigation on a number of salts, Pankratov (1939) and Frumkin and Pankratov (1939) obtained similar results. In order to be able to compare surface pressures at the same thermodynamic potential of film-forming compound, they measured also the increase in the equilibrium spreading

* Chemistry Department, University of Melbourne.

pressure (E.S.P.), i.e. the pressure exerted by a film in equilibrium with excess bulk phase, due to salts.

Donnison and Heymann (1946) investigated the E.S.P. of oleic acid and of ethyl sebacate on concentrated solutions of a number of uni-univalent salts over a limited range of concentrations using a ring method. Their results indicated a linear relationship between E.S.P. and activity of salt in the subphase.

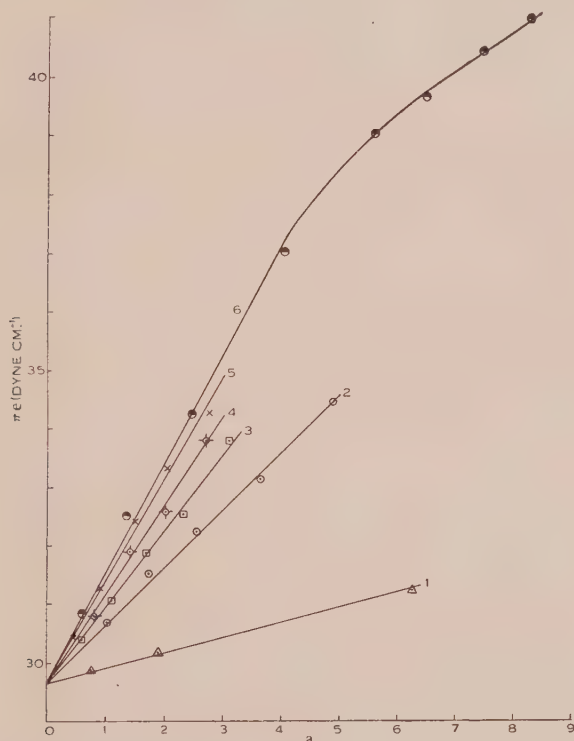


Fig. 1.—E.S.P.-activity curves for oleic acid on solutions of
(1) LiCl; (2) NaCl; (3) NH_4Cl ; (4) KCl; (5) KBr;
(6) KCNS; at 20 °C.

The present investigation was undertaken to confirm and supplement the work of Donnison and Heymann (1946). The E.S.P. and π - A curves of oleic acid, on solutions of a range of salts and non-ionic compounds, have been measured at concentrations of up to saturation.

II. EXPERIMENTAL

The surface balance used was a modification of the Langmuir-Adam horizontal pull type with a single torsion wire and was built in the Colloid Science Department, Cambridge. The trough was of Pyrex glass and the gaps between the trough and the mica float were closed with 0.0005 in. thick platinum strips in the simple form suggested by Adam (1941, p. 31). The apparatus etc. was kept clean with sulphuric acid-dichromate cleaning mixture.

Since equilibrium values are hard to obtain with solids in measurements of E.S.P., due either to the sluggishness of spreading from crystals (Pankratov 1939) or the presence of impurities (Carey and Rideal 1925), a liquid film-forming substance was chosen. Because of the type of balance used, the film had to be insoluble and oleic acid was therefore chosen; moreover it was convenient because of its large E.S.P. and its use in earlier work (Donnison and Heymann 1946). The oleic acid used was a pure Schering-Kahlbaum product, the E.S.P. of which was independent of the amount on the surface, this being the criterion of purity of Carey and Rideal (1925).

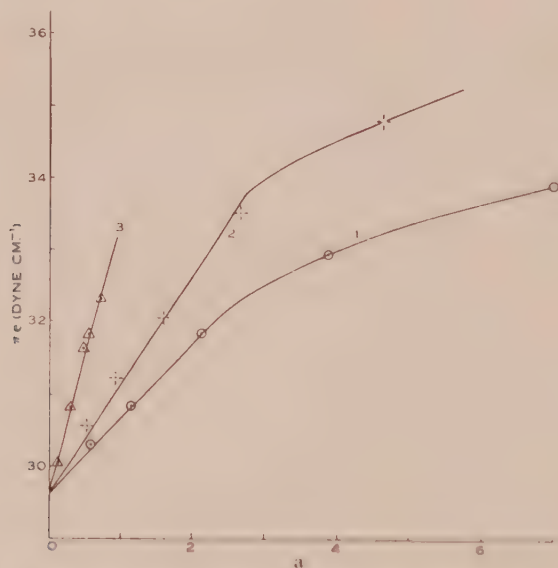


Fig. 2.—E.S.P.-activity curves for oleic acid on solutions of (1) MgCl_2 ; (2) CaCl_2 ; (3) BaCl_2 ; at 20°C .

All salts used were of A.R. quality. Organic impurities were eliminated from the sodium and potassium halides, barium and calcium chlorides, and sodium and magnesium sulphates by heating to a red heat (Jones and Ray 1937; Pankratov 1939). The solutions of ignited barium and calcium chloride were alkaline and had to be neutralized. Solutions of other compounds which could not be heated were allowed to stand with a large surface exposed so that impurities could concentrate in the surface and be removed.

The solutions were made as concentrated as possible initially and then successively diluted. Before use they were filtered through a fine sintered glass filter and extracted with light petroleum. Fresh solutions of sucrose were used to minimize effects of inversion.

Twice distilled water was used in the subphase which was always kept 0.01N with respect to hydrochloric acid. Carey and Rideal (1925) recommend this to avoid soap formation which renders results unrepeatable. Sebba and Briscoe (1940), however, ascribe these effects to solubility of the oleic acid.

Initially, the method of working was essentially that of Adam and Jessop (1926). However, unsatisfactory results were obtained, the E.S.P. of oleic acid being some 10 per cent. higher than the results of previous investigators. Further, the E.S.P. increased slowly with time. Purification of the paraffin wax, by repeated treatments with concentrated sulphuric acid and sodium hydroxide, enabled satisfactory results to be obtained although the wax deteriorated again rapidly. Moreover, waxes from various sources behaved similarly. Adam, Askew, and Pankhurst (1939) have shown that adsorption of a non-polar or slightly polar substance disrupts the coherence of a film and causes an increase in the surface pressure. The effect of a non-polar oil mixed with a film is greatest in the liquid-expanded state (Harkins and Myers 1936*a*, 1936*b*). The impurity in the paraffin wax may be such a very slow spreading compound resulting, perhaps, from oxidation.

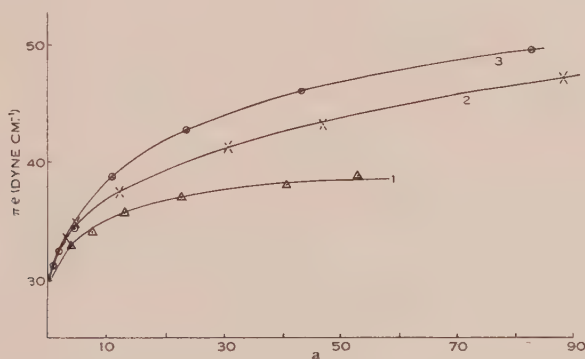


Fig. 3.—E.S.P.-activity curves for oleic acid on solutions of (1) MgCl_2 ; (2) CaCl_2 ; (3) LiCl^* ; at 20°C .

* Note.—For curve 3 multiply abscissa scale by 10.

Previously, Langmuir and Schaefer (1937) had failed to purify paraffin wax and had replaced it with ferric stearate. This technique was adopted and ferric stearate was used to coat the trough, glass barriers, and float, and also to cement the platinum strips. Steady readings were then obtained.

(a) *Equilibrium Spreading Pressures*

A drop of oleic acid was placed on the cleaned surface of the subphase with a flamed platinum wire. A steady reading was obtained usually within 20 seconds of spreading. The value of the E.S.P. was corrected to 20°C . using the temperature coefficient of $0.071 \text{ dyne/cm.}/^\circ\text{C}$. (Carey and Rideal 1925).

On increasing the area available to the film, the film was rapidly replenished from the lenses of excess oleic acid and the E.S.P. resumed usually within 10 seconds. On compression, however, the resumption of equilibrium often took several minutes. On salt solutions, the attainment of equilibrium was slower and permanent increases of up to 1 dyne cm.^{-1} were found on compression of the film, the original reading being resumed only on increasing the area. With

increase in area on salt solutions, the original E.S.P. was regained slowly. This afforded an argument against the use of the ring method of measurement for these systems.

(b) Force-Area Curves

Oleic acid was placed on the cleaned surface to give a definite area per molecule by delivering a standard solution in purified light petroleum from an Agla micrometer syringe. A steady pressure reading was obtained within 30–45 seconds after each movement of the barrier and π - A curves plotted. Measurements with the barrier close to the float were avoided (Carey and Rideal 1925).

Since films exhibit considerable hysteresis on re-expansion after compression (Harkins and Fischer 1933), π - A curves were taken only on the forward run. The rate of compression of the film further affects results (Nutting and Harkins 1939) but no detailed study has been made of this factor. In expanded films, as with oleic acid, the effect is small, but the molecular area at which collapse occurs will be affected.

(c) Accuracy of Results

All values of the E.S.P. are the mean of three readings not differing by more than 0.1 dyne cm.⁻¹. The E.S.P. of oleic acid on 0.01N hydrochloric acid at 20 °C. was found to be 29.7 dyne cm.⁻¹ which compares well with the value of 29.4 dyne cm.⁻¹ of Carey and Rideal (1925), 29.5 dyne cm.⁻¹ of Blodgett (1935), and 29.6 of Donnison and Heymann (1946).

An analysis of errors after Harkins and Anderson (1937) shows that with the surface balance used the total systematic error in measurements of surface pressure may be as great as ± 3 per cent., whilst a random error of ± 0.08 dyne cm.⁻¹ is likely. Measurements did not vary by more than ± 0.25 per cent. on varying the water level which is within the estimated error. The area per molecule of oleic acid was known to about ± 0.5 per cent. and the concentration of solutions to within 0.2 per cent. Dilution due to surface adsorption of atmospheric water could not be detected.

Quite a large uncertainty is introduced in the choice of activity coefficients.* However, the relative accuracy within any one worker's results is usually of the order of 0.1 per cent.

* The sources of the activity coefficients were:

KCl, NaCl, NH₄Cl, and Sucrose: Landolt-Bornstein, *Physikalisch-Chemische Tabellen*.

KBr (also other alkali halides): Robinson, R. A. (1935).—*J. Amer. Chem. Soc.* **57**: 1161.

KCNS: Robinson, R. A. (1940).—*J. Amer. Chem. Soc.* **62**: 3131.

LiCl: Robinson, R. A. (1945).—*Trans. Faraday Soc.* **41**: 756.

Na₂SO₄: Robinson, R. A., Wilson, J. M., and Stokes, R. H. (1941).—*J. Amer. Chem. Soc.* **63**: 1012.

MgSO₄: Robinson, R. A., and Jones, R. S. (1936).—*J. Amer. Chem. Soc.* **58**: 959.

MgCl₂: Stokes, R. H. (1945).—*Trans. Faraday Soc.* **41**: 642.

BaCl₂: Robinson, R. A. (1940).—*Trans. Faraday Soc.* **36**: 735.

CaCl₂: Stokes, R. H. (1945).—*Trans. Faraday Soc.* **41**: 642.

III. RESULTS AND DISCUSSION

(a) *Equilibrium Spreading Pressures*

The results of measurements of the E.S.P. of oleic acid (corrected to 20 °C.) with a range of salts in the subphase are given in Figures 1-4, which show the E.S.P., πe , plotted against the activity, a , of the salt in the subphase.

Figure 1 shows the results for a number of univalent salts. An approximate linear relationship is revealed, confirming the results of Donnison and Heymann (1946). However, as shown in the curves for potassium thiocyanate and lithium chloride, the linear relationship does not extend beyond an activity of 5-6, after which point a smooth curve is obtained. In Figure 3 is a separate plot

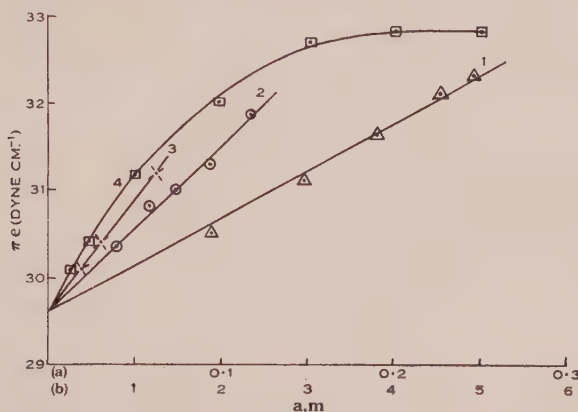


Fig. 4.—E.S.P.-activity curves for oleic acid on solutions of (1) Na_2SO_4 —abscissa scale (a); (2) MgSO_4 —abscissa scale (a); (3) sucrose—abscissa scale (b); at 20 °C. Also (4) E.S.P.-molality curve for oleic acid on sucrose solutions at 20 °C.—abscissa scale (b).

on a smaller scale of the whole range studied for lithium chloride. The remarkable increase in E.S.P. on the most concentrated (about 16.5 molar) lithium chloride solutions is worthy of note.

The results for magnesium, calcium, and barium chlorides are shown in Figure 2. Again a linear relationship between E.S.P. and activity is maintained up to about $a=5$, the limit being somewhat lower for magnesium chloride. The range for magnesium and calcium chlorides is extended in Figure 3, smooth curves being obtained after the initial linear portion. It may be noted that the most concentrated calcium chloride solution (about 6 molar) produces an increase in the E.S.P. almost as great as the much more concentrated lithium chloride.

Figure 4, which illustrates sodium and magnesium sulphates, shows a linear πe - a curve over the whole concentration range.

So far as the activity data extend, sucrose shows a linear relationship between E.S.P. and activity. This is illustrated also in Figure 4, where a plot of πe against molal concentration is also shown over a greater concentration range. The E.S.P. appears to reach a limiting value with increase in sucrose concentration.

The effectiveness of salts in increasing the E.S.P. may be compared as follows:

If the E.S.P. of oleic acid on water is πe° and that on a salt solution of activity a is πe , then, provided $a > 5$ or 6,

$$\pi e = \pi e^\circ + \beta a.$$

The values of β found experimentally are given for comparison in Table 1.

TABLE 1
EXPERIMENTAL VALUES OF β
(Temperature 20 °C.)

Salt	β	Salt	β
LiCl	0.25	KCl	1.52
NaCl	1.00	KBr	1.76
NH ₄ Cl	1.30	KCNS	1.82
KCl	1.52		
MgCl ₂	1.03	Na ₂ SO ₄	10.9
CaCl ₂	1.55	MgSO ₄	18.4
BaCl ₂	3.82		
Sucrose	1.21		

It will be seen on comparing ions of the same valence type that the order of effectiveness of cations and of anions is $\text{Li}^+ < \text{Na}^+ < \text{NH}_4^+ < \text{K}^+$, $\text{Mg}^{++} < \text{Ca}^{++} < \text{Ba}^{++}$, and $\text{Cl}^- < \text{Br}^- < \text{CNS}^-$. It should be noted here that the influence of the fluoride and iodide ions could not be studied since, in the one case, oleic acid is attacked by traces of hydrogen fluoride liberated in acid solution and, in the other, iodine is liberated beneath the oleic acid film possibly through the oleic acid acting as an oxygen carrier (Donnison and Heymann 1946).

That the above order of increasing effect on the E.S.P. is also the order of decreasing energy of hydration of the ions is evident from Table 2 (Butler 1940, p. 38).

TABLE 2
ENERGY OF HYDRATION OF VARIOUS IONS (KCAL./MOLE)*

Li ⁺	144	Cl ⁻	65	Mg ⁺⁺	498
Na ⁺	117	Br ⁻	53	Ca ⁺⁺	421
K ⁺	95			Ba ⁺⁺	347
NH ₄ ⁺	97				

* Butler (1940).

Comparison with the lyotropic series (the salting out effect on lyophilic colloids by electrolytes) is also instructive. This effect is due to the tendency of the ions to become hydrated. The order of cations is



and for anions



Thus the general rule both for the anions and for the cations of one valence type is that the larger the hydration energy of the ion the weaker is its effect on the E.S.P.

As well as the hydration effect, however, there appears also to be a charge effect. This is probably best shown in the case of the sulphate ion which is the

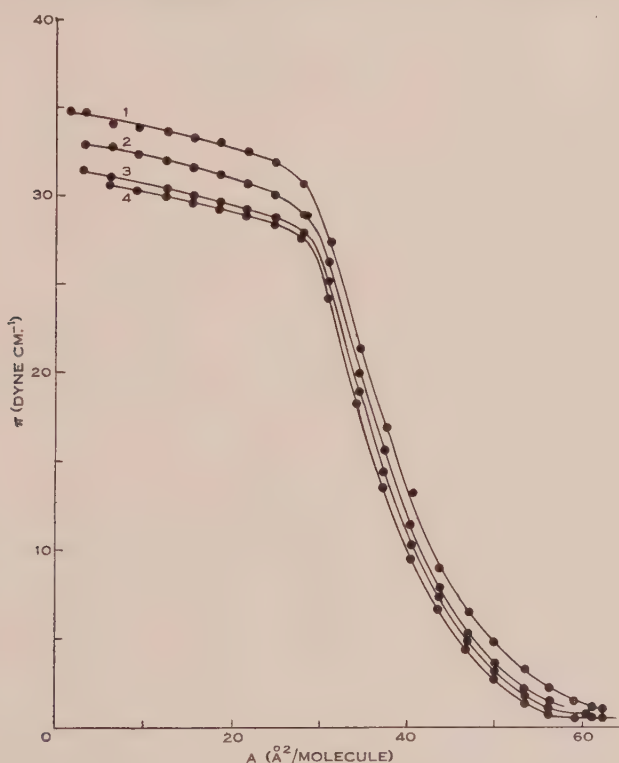


Fig. 5.— π - A curves for oleic acid on

(1) NaCl	5.5m	19 °C.
(2) NaCl	3.5m	19 °C.
(3) NaCl	0.5m	16 °C.
(4) HCl	0.01N	19.5 °C.

most strongly hydrated of the anions investigated and would be expected to have the least influence on the E.S.P. However, the values of β for both sodium sulphate and magnesium sulphate are far larger than for the salts of any other anion. That a charge effect counteracting the hydration effect exists is suggested also by comparing the univalent chlorides with the bivalent chlorides. However, the charge effect is more marked with anions than with cations and is greatest in the case of the bi-bivalent salt, magnesium sulphate.

Of special interest in the consideration of a charge effect is the fact that sucrose causes an appreciable increase in the E.S.P. Sucrose increases the

surface tension of water and hence is negatively adsorbed. This adsorption will be altered in the presence of the film (sucrose has six polar groups to the molecule) to produce an increase in the work of adsorption on the surface and a rise in the E.S.P. Since sucrose is a non-electrolyte and there can be no ionic forces, the presence of charged particles in the surface is not essential to increase the E.S.P.

Thus, to explain the E.S.P. results for salt solutions, there are envisaged two effects :

(i) A hydration effect, as revealed by comparison of the anions or cations which are common to the salts within any one valency group. This is due to

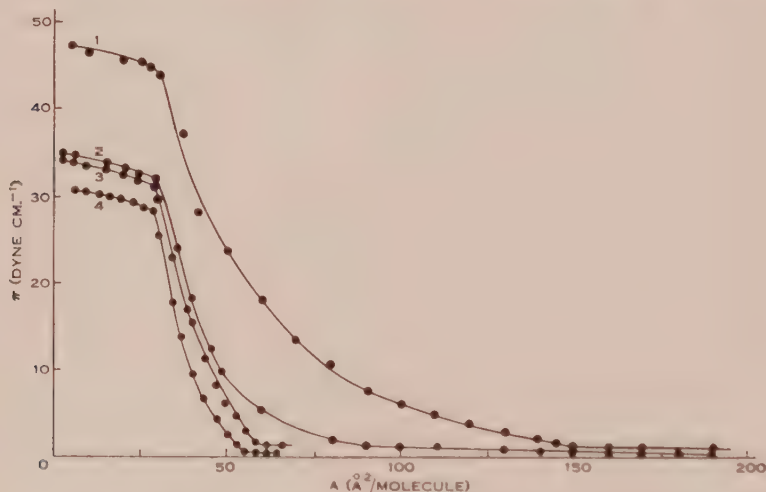


Fig. 6.— π - A curves for oleic acid on

(1) LiCl	16.6m	20.5 °C.
(2) LiCl	8.0m	21 °C.
(3) KCl	4.5m	21 °C.
(4) HCl	0.01N	19.5 °C.

the fact that the less strongly hydrated ions are adsorbed in the surface more strongly than ions of greater hydration.

(ii) A charge effect, the magnitude of which depends on the valence type of the salt, anions having a greater effect than cations. The charge effect masks and counteracts the differences expected due to the hydration of ions of different valency. The dependence of the E.S.P. on the charge of the ions is probably due to greater increase in the adsorption of ions in the presence of film than is obtained with ions of lesser charge, since the interaction between ions and the polar groups of the film will be stronger, the greater the charge of the ion. Since anions are already nearer the surface, on the average, than cations (Frumkin 1924), they would be expected to have the larger effect.

(b) Force-Area Curves

The π - A curve of oleic acid on 0.01N hydrochloric acid is shown in Figure 5, curve 4. Both the shape of the curve, which is typical of liquid expanded type,

and its position are in good agreement with the results of other investigators (Langmuir 1917; Adam and Jessop 1926; Harkins and Fischer 1933; Der-

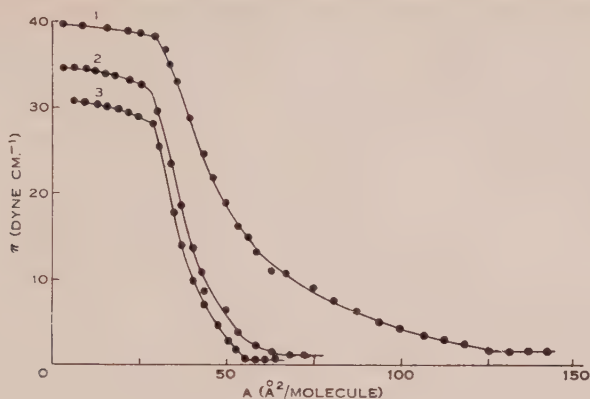


Fig. 7.— π - A curves for oleic acid on

(1) KCNS	17.0m	20 °C.
(2) KBr	4.5m	18.5 °C.
(3) HCl	0.01N	19.5 °C.

vichian 1935) although the conditions are usually not strictly comparable. The limiting area of the liquid expanded phase is 55–56 Å² (cf. 55 Å², Adam and

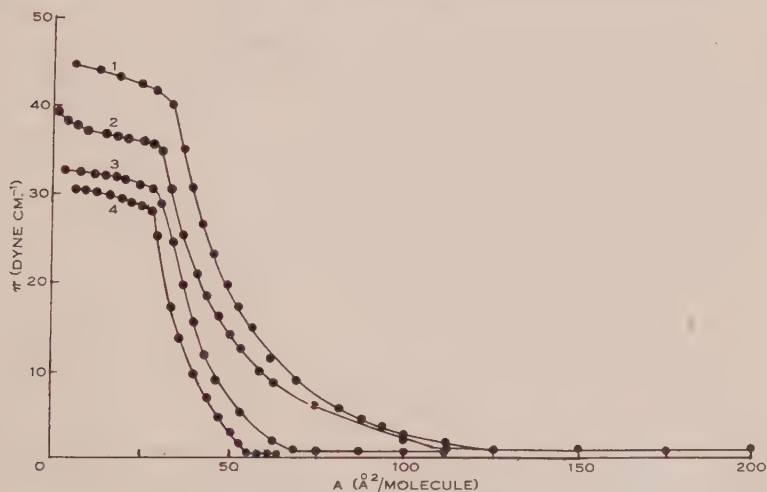


Fig. 8.— π - A curves for oleic acid on

(1) CaCl ₂	8.3m	21 °C.
(2) BaCl ₂	1.70m	20 °C.
(3) MgCl ₂	4.8m	20.5 °C.
(4) HCl	0.01N	19.5 °C.

Jessop 1926) and the point of collapse of the film is at an area of 29–30 Å² (cf. 28–30 Å², Dervichian 1937).

In Figure 5 are compared also the π - A curves of oleic acid on sodium chloride solutions at three concentrations. It will be seen that the curves maintain their

characteristic shape and that, on the salt solutions, there is an increase of area at constant surface pressure and an increase of surface pressure at constant area per molecule. This effect is greater, the more concentrated the salt. This confirms the results of Pankratov (1939), who found a similar behaviour with films of cetyl alcohol and of ethyl palmitate.

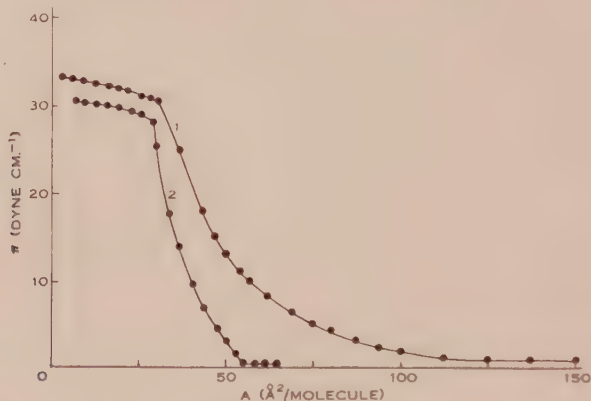


Fig. 9.— π - A curves for oleic acid on
(1) Sucrose 5.0m 20 °C.
(2) HCl 0.01N 19.5 °C.

Figures 6-9 show the π - A curves for various salts and for sucrose as indicated in each figure. The π - A curves on sodium sulphate and magnesium sulphate solutions have not been included. These are almost identical with the 0.01N hydrochloric acid subphase curve except for a shift to larger areas, being very similar to curves 1 and 2 in Figure 5. In Table 3 the reference points are given.

TABLE 3
REFERENCE POINTS FOR π - A CURVES ON Na_2SO_4 AND MgSO_4 SOLUTIONS

	Area per Molecule (\AA^2) as :			Limiting Area (\AA^2)
	π (dyne cm^{-1}) = 30	20	5	
Na_2SO_4 (1.33m, 20 °C.) ..	25	35	49	62
MgSO_4 (2.18m, 20.2 °C.) ..	26	37	53	75

The π - A curves on all the salt solutions are still of the liquid expanded type with a shift to larger areas. All the curves show the point of collapse at the same area as in the film on dilute hydrochloric acid, i.e. at a little under 30 \AA^2 per molecule. This again is in agreement with the results of Pankratov which indicate that for ethyl palmitate there is a transition from the expanded state

to an intermediate type of film at an area of about 50 \AA^2 per molecule which is not changed by a salt in the subphase. However, oleic acid is so far above its expansion temperature range that it maintains its liquid expanded form right up to collapse of the film. The area at which collapse of the film commences indicates that, at this area, the film begins to lose its monomolecular character and the first appearance of a three-dimensional phase occurs. The fact that the area of collapse is the same irrespective of the presence of salt in the subphase may suggest that the salt does not change the geometry of the film and makes it highly improbable that chemical effects occur in the films. However, the pressure at the point of collapse does not correspond to the E.S.P. This is reached only on further considerable reduction of the area.

A comparison of the various π - A curves shows further that, in the liquid expanded part, the increase in the area due to the salt is rather small at large values of π , whereas, at lower values of π when the film is more dilute, the increase in area is generally larger. Alternatively, in the more dilute film, a given reduction in area produces a smaller increase in surface pressure than the same reduction in the smaller area range. This would be expected on the basis of an increase in the repulsive terms in the system of forces between the molecules. The further apart the film molecules are, the less significant such forces will be in determining the film pressure.

The conditions of Frumkin and Pankratov (1939) whereby they deduce an increase in the work of adsorption of the film on the surface are fulfilled by our curves also. That is, comparing any two curves on, say, solutions 1 and 2, both the E.S.P. and the force at constant area are greater on solution 1 than on solution 2. Hence, by the same analysis, it may be deduced that the work of adsorption on solution 1 is greater than on solution 2.

Thus the increase of the E.S.P. and of the spreading pressure at constant area can be explained by visualizing the same two effects put forward by Frumkin and Pankratov and used by Donnison and Heymann, namely :

- (i) An increase of the attraction between organic molecules and the substrate.
- (ii) An increase of the repulsive terms of the system of forces between the molecules.

A useful summary of the forces giving rise to these effects has been given by Harkins (1944). Changes in configuration within the carboxyl group of the oleic acid on the model of Adam and Harding (1931) may also contribute. The charge and hydration effects probably stem largely from differences in adsorption of ions and changes in adsorption due to the presence of film.

IV. ACKNOWLEDGMENTS

During the course of this work one of us (A.R.G.) was in receipt of a scholarship from the Daffyd Lewis Trust. Thanks are due to Dr. A. S. Buchanan, Chemistry Department, University of Melbourne, for checking the manuscript.

V. REFERENCES

- ADAM, N. K. (1941).—"The Physics and Chemistry of Surfaces." 3rd Ed. (Oxford Univ. Press.)
- ADAM, N. K., ASKEW, F. A., and PANKHURST, K. J. A. (1939).—*Proc. Roy. Soc. A* **170** : 485.
- ADAM, N. K., and HARDING, J. B. (1931).—*Proc. Roy. Soc. A* **130** : 284.
- ADAM, N. K., and JESSOP, G. (1926).—*Proc. Roy. Soc. A* **110** : 423.
- BLODGETT, K. (1935).—*J. Amer. Chem. Soc.* **57** : 1008.
- BUTLER, J. A. V. (1940).—"Electrocapilarity." (Methuen & Co. Ltd. : London.)
- CAREY, A., and RIDEAL, E. K. (1925).—*Proc. Roy. Soc. A* **109** : 301, 318.
- DERVICHIAN, D. G. (1935).—*J. Phys. Radium* **6** : 429.
- DERVICHIAN, D. G. (1937).—*Ann. Phys.* **8** : 436.
- DONNISON, JOAN A., and HEYMANN, E. (1946).—*Trans. Faraday Soc.* **42** : 1.
- FRUMKIN, A. (1924).—*Z. Phys. Chem.* **109** : 34.
- FRUMKIN, A., and PANKRATOV, A. (1939).—*Acta Physicochim. U.R.S.S.* **10** : 55.
- GLASSTONE, S. (1943).—"Text-Book of Physical Chemistry." (Van Nostrand Co. : New York.)
- HARKINS, W. D. (1944). "Colloid Chemistry, Theory and Methods." Vol. 5, p. 38. (Reinhold Publ. Corp. : New York.)
- HARKINS, W. D., and ANDERSON, T. F. (1937).—*J. Amer. Chem. Soc.* **59** : 2189.
- HARKINS, W. D., and FISCHER, E. K. (1933).—*J. Chem. Phys.* **1** : 852.
- HARKINS, W. D., and MYERS, R. J. (1936a).—*J. Amer. Chem. Soc.* **58** : 1817.
- HARKINS, W. D., and MYERS, R. J. (1936b).—*J. Phys. Chem.* **40** : 959.
- HARKINS, W. D., and MORGAN, J. W. (1925).—*Proc. Nat. Acad. Sci.* **11** : 637.
- JONES, G., and RAY, W. A. (1937).—*J. Amer. Chem. Soc.* **59** : 187.
- LANGMUIR, I. (1917).—*J. Amer. Chem. Soc.* **39** : 1848.
- LANGMUIR, I., and SCHAEFER, V. J. (1936).—*J. Amer. Chem. Soc.* **58** : 284.
- LANGMUIR, I., and SCHAEFER, V. J. (1937).—*J. Amer. Chem. Soc.* **59** : 2400.
- MITCHELL, J. S., RIDEAL, E. K., and SCHULMAN, J. H. (1937).—*Nature* **139** : 625.
- MYERS, R. J., and HARKINS, W. D. (1937).—*Nature* **139** : 367.
- NUTTING, G. C., and HARKINS, W. D. (1939).—*J. Amer. Chem. Soc.* **61** : 2040.
- PANKRATOV, A. (1939).—*Acta Physicochim. U.R.S.S.* **10** : 45.
- SEBBA, F., and BRISCOE, H. V. A. (1940).—*J. Chem. Soc.* **1940** : 106.

THE MAGNETIC PROPERTIES AND CHEMICAL STRUCTURES OF SOLID HAEMINS

By W. A. RAWLINSON* and P. B. SCUTT†

[*Manuscript received September 26, 1951*]

Summary

The magnetic properties of protoporphyrin and a representative group of its iron derivatives have been studied in order to gain more detail of their structural characteristics. The greater diamagnetism of protoporphyrin than that expected on the basis of Pascal's additivity law indicates an increased stability of pyrrole units when linked in resonating porphin form. The majority of iron-porphyrin complexes possess ionically bound metal when crystallized from non-aqueous liquids. In the presence of water, the lower values of magnetic moment reveal that iron tends to form covalent bonds. The iron bonds of various haematin compounds are compared with those of simple iron hydroxides and ferritin. There is strong evidence to support the assumption that haematin in dilute alkali possesses covalent iron bonds involving only one *d* eigenfunction. In concentrated alkali the magnetic moment of haematin (2.4) indicates that six covalent iron bonds should be directed to the corners of either an octahedron or a trigonal prism, but the steric restrictions imposed by the planar porphyrin nucleus and the covalent bond angle of oxygen make it difficult to assign such specific bond directions to the metal.

I. INTRODUCTION

Measurements of magnetic susceptibilities have been used in the investigation of many biologically important haem compounds, the significance of such measurements being that in these iron-porphyrin compounds the paramagnetic contributions may yield information on the types of chemical bonding to the iron atoms. In particular, such measurements have been made on haemoglobin and its derivatives by workers of Pauling's school, and on cytochrome *c*, peroxidases, and catalases by Theorell and his co-workers. From the results of these workers and others (see review by Hartree 1946), it has been established whether, in such derivatives, the iron bonds were predominantly ionic or covalent. In these compounds the haem prosthetic groups are bound to specific proteins, and in consequence the iron contents are rather low and accurate measurements of magnetic moments difficult. The measurements presented here have been made on the simpler types of haem compounds in order to reveal the significance of differences of magnetic susceptibilities. Hartree has referred to the need for such studies, and in his review many of the results presented below have been quoted. Selwood (1943) has made a valuable review of the magnetochemistry of iron complexes, although he has made an unfortunate error in the case of iron-porphyrins by assuming that haematin and methaemoglobin are identical.

* Department of Biochemistry, University of Melbourne.

† Institute of Agriculture, University of Western Australia, Nedlands, W.A.

The only published values of magnetic susceptibilities for solid compounds appear to be those of Haurowitz and Kittel (1933) for haemin and haematin, Cambi and Szego (1934) for haemin at room temperature and at lower temperatures, and Pauling and Coryell (1936) for haemin. The compounds measured by us include those in which (a) the chloride ion of haemin is replaced by other anions, (b) the peripheral substituents of the porphyrin nucleus are varied, and (c) the porphyrin ring itself is altered by replacement of one methene bridge by nitrogen. The values obtained for the magnetic moments indicate that, although in most cases the bonding of the iron still appears essentially ionic, there are

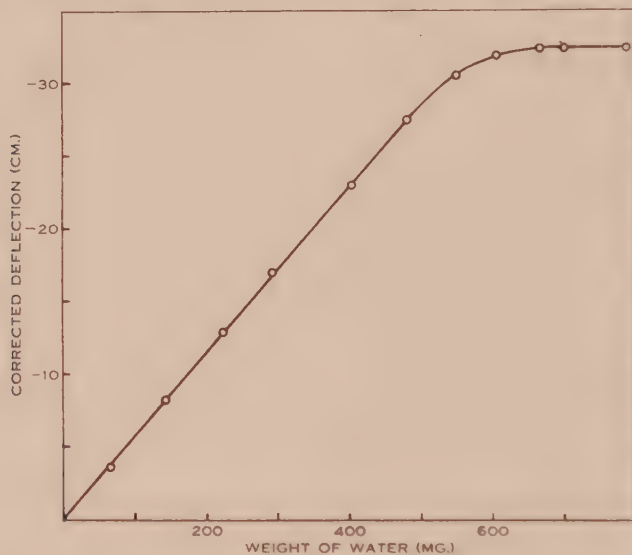


Fig. 1.—The relationship between tube deflection and weight of water at various levels in the magnetic field.

significant differences of magnetic moment. The variation of magnetic moment of haematin (ferriporphyrin hydroxide) components with the conditions of preparation shows some points of interest which are compared with those of other iron compounds, ferritin in particular.

II. MEASUREMENT OF MAGNETIC PROPERTIES

The measurements were made on a Curie-Cheneveau magnetic balance of the type described by Gray, Clow, and Cruickshank (1936). The construction and use of the particular instrument have been reported earlier by Rawlinson (1940).

The actual technique of measurement has been modified so as to obtain maximum accuracy when using small samples of paramagnetic solids. The reason for this modification is evident from examination of the curve shown in Figure 1, in which is plotted the tube deflection (corrected for air displaced and dissolved air, according to equation (1)) against weight of water, when the tube bottom was placed level with the bottom edges of the pole pieces of the magnet.

The response is strictly linear up to a weight of 480 mg. of water, corresponding to a water level about 4 mm. below the top edges of the pole pieces. Beyond this point the deflection falls away from proportionality to reach a maximum (corresponding to the "active volume" of Gray, Clow, and Cruickshank 1936), so that further additions of water produce no increase of deflection. In all measurements and standardizations, therefore, the level of material in the tube was kept more than 4 mm. below the top of the pole pieces.

The following equation was used to correct the actual deflections for plotting the values in Figure 1.

$$\left. \begin{aligned} D^{\circ} &= (D' - D'')(1 - Ad'/\chi' + \alpha A^{\circ}d'/\chi'), \\ &= 1.0381(D' - D''), \end{aligned} \right\} \dots\dots\dots (1)$$

where D° , D' , and D'' represent corrected deflection, deflection of tube plus water, and empty tube respectively; A and A° , the volume susceptibilities of air and dissolved air, respectively, $+0.0294 \times 10^{-6}$ and $+0.0593 \times 10^{-6}$; d' the density of water; α the solubility of air in water; and χ' the mass susceptibility of water.

The standard substance used was glass-distilled water which was taken to have a mass susceptibility, $\chi' = -0.720 \times 10^{-6}$. At least one standardization was made each day. The daily variation was slight but over a period of 2 years for which a single fibre was in use there was a definite increase in the deflection for a given mass of water. As shown in equation (1) the correction for air dissolved in water is not significant so it has not been included in our routine calculations below.

The mass susceptibility (χ) of material under test was calculated from the following equation of Gray, Clow, and Cruickshank (1936), which includes the correction for the air displaced by the sample:

$$\chi = \frac{m'}{m} \cdot \frac{D - D''}{D' - D''} \cdot (\chi' - A/d') + A/d, \dots\dots\dots (2)$$

where m and m' are the weights of substance and water respectively; D , D' , and D'' are the actual deflections of tube plus substance, tube plus water, and empty tube respectively; χ' is the mass susceptibility of water; A is the volume susceptibility of air; and d and d' are the densities of substance and water respectively.

The density values used by us are the absolute densities and not the "apparent densities" of packed materials as used by Gray, Clow, and Cruickshank (1936). By flotation in chloroform-benzene mixtures solid protohaemin was found to have a density of 1.43 g.cm.^{-3} . A considerable variation in this figure is not significant, affecting only the fourth place in susceptibility values, and it has been assumed that other compounds studied have densities of the same order.

The molar susceptibility (χ_{mol}) was calculated from the actual Fe content of each sample, on the assumption that each pigment molecule contained one Fe atom.

$$\chi_{mol} = 5585\chi/\text{per cent. Fe.} \dots\dots\dots (3)$$

Iron estimations were carried out by the method of Ramsay (1944) using samples of about 10 mg. In a few cases the method of Drabkin (1941) was used when small amounts of sample only were available. The results referred to in Table 2 are mean values.

In order to obtain a true measure of the paramagnetic susceptibility of the Fe present, it was necessary to correct for the diamagnetism of the rest of the molecular structure in the case of each compound. For this purpose a sample of recrystallized protoporphyrin No. 9 was measured and gave values

$$\chi_{mass} = -0.719 \times 10^{-6} \text{ and } -0.713 \times 10^{-6},$$

giving a mean molar susceptibility of -399.9×10^{-6} . This figure was used as the basis for diamagnetic corrections. For other groups coordinated with the iron, and for changes in the porphyrin molecule the normal Pascal constants and ionic susceptibility values were used.

From the corrected molar susceptibilities the magnetic moments (μ) were calculated according to the equation

$$\mu = 2.839 \sqrt{\chi_{mol} \cdot T}. \quad (4)$$

The use of this equation assumes the validity of the Curie-Weiss law,

$$\chi_{mol} = C_M / (T - \theta), \quad (5)$$

and that the Curie temperature (θ) is absolute zero. This assumption was verified by the work of Cambi and Szego (1934) on haemin, and has been assumed to hold for other haem compounds. The relationship of magnetic moment (μ) to the number of unpaired electrons (n) per iron atom is given by equation (6) (see Table 1).

$$\mu = \sqrt{n(n+2)}. \quad (6)$$

TABLE 1
THE RELATIONSHIP BETWEEN MAGNETIC PROPERTY AND NUMBER OF
UNPAIRED ELECTRONS

n	1	3	5
$\chi_M \times 10^6$ at 20 °C.	1270	6350	14820
μ_B	1.73	3.87	5.92

III. PREPARATION OF MATERIALS

A. 1. *Protoporphyrin No. 9* ($C_{34}H_{34}O_4N_4$) was prepared by the method of Fischer and Putzer (1926). After recrystallization from pyridine-chloroform the product was dissolved in pyridine and crystallized by adding light petroleum (40-60 °C.). Final crystallization was made by saturating separately with the pigment 1 vol. pyridine and 10 vols. glacial acetic acid, each with careful boiling. By running the hot filtrate from the acid solution into the filtered pyridine solution, large crystals of protoporphyrin separated within a few minutes. After standing overnight, the crystals were filtered, washed with glacial acetic acid, and dried *in vacuo* over KOH.

- B. 1. *Protohaemin chloride* ($C_{34}H_{32}O_4N_4 \cdot FeCl$) was prepared by the modified method of Schalfhiev and recrystallized according to Hogness *et al.* (1937). 2. *Protohaemin chloride dimethyl ester* ($C_{36}H_{36}O_4N_4 \cdot FeCl$) was prepared by the method of Küster (1927). 3. *Protohaemin bromide* ($C_{34}H_{32}O_4N_4 \cdot FeBr$) was prepared in the same manner as protohaemin chloride using potassium bromide instead of sodium chloride (Fischer, Treibs, and Zeile 1930). 4. *Protohaemin formate* ($C_{34}H_{32}O_4N_4 \cdot Fe \cdot O \cdot CO \cdot H$) was prepared by the method of Hamsik (1929a). 5. *Protohaemin acetate* ($C_{34}H_{32}O_4N_4 \cdot Fe \cdot O \cdot CO \cdot CH_3$) was prepared according to Hamsik (1930). 6. *Protohaemin propionate* ($C_{34}H_{32}O_4N_4 \cdot Fe \cdot O \cdot CO \cdot C_2H_5$) was prepared in similar manner to protohaemin acetate. This compound has not been reported previously and is of interest on account of the presence of propionic acid groups in protoporphyrin derivatives. 7. *Mesohaemin chloride* ($C_{34}H_{36}O_4N_4 \cdot FeCl$) was prepared from mesoporphyrin by the usual method and recrystallized from glacial acetic acid. 8. *Deuterohaemin chloride* ($C_{30}H_{23}O_4N_4 \cdot FeCl$) was prepared by the method of Fischer and Hummel (1929). 9. *Azaahaemin chloride* ($C_{33}H_{31}O_4N_5 \cdot FeCl$) was prepared by the method of Lemberg (1943). 10, 11. *Haematin*, pseudocrystalline ($C_{34}H_{32}O_4N_4 \cdot FeOH$) was prepared by two methods described by Hamsik (1936), (a) by hydrolysis of protohaemin acetate, (b) by decomposition of pyridine haematin. 12. *Haematin half-anhydride* ($C_{68}H_{64}O_9N_8 \cdot Fe_2$) was prepared from protohaemin acetate by the method of Hamsik (1936). 13. *Haematin anhydride* ($C_{34}H_{31}O_4N_4 \cdot Fe$) was prepared by the method of Hamsik (1924). 14. *Haematin barium salt* ($C_{34}H_{30}O_4N_4 \cdot Ba \cdot FeOH$) was prepared according to Hamsik (1923-4). 15. *Haematin*, amorphous ($C_{34}H_{32}O_4N_4 \cdot FeOH$) was prepared by dissolving protohaemin chloride in 0.2M KOH, and precipitating with dilute acetic acid. The precipitate was dissolved and reprecipitated as above and then washed with water and centrifuged repeatedly until the supernatant contained colloidal haematin. The precipitate was then dried *in vacuo* over KOH. 16. *Haematin dipotassium salt* ($C_{34}H_{30}O_4N_4 \cdot K_2 \cdot FeOH$) was prepared by dissolving protohaemin chloride in 0.2M KOH, and then precipitating by adjustment to final 9.5M KOH concentration. 17, 18. *Haematin dipotassium salts* were prepared by dissolving protohaemin chloride or freshly dried, amorphous haematin in 0.1M KOH and precipitation from final KOH of 5M and 7.5M concentrations respectively. The precipitates were washed well with ethanol and dried *in vacuo* over KOH and H_2SO_4 . 19. *Haematin tripotassium salt* ($C_{34}H_{30}O_4N_4 \cdot K_3 \cdot FeOK$) was prepared according to Hamsik (1929b).

IV. PROTOPORPHYRIN AND DIAMAGNETIC CORRECTIONS

The comparison of magnetic susceptibilities of the iron atoms present in the haem pigments studied was made after correcting for the diamagnetic contribution of the complete molecule. As the majority of compounds investigated were related to natural protoporphyrin (No. 9), it was used as the basis for corrections. The molar susceptibility of carefully purified protoporphyrin was

found to be -400×10^{-6} . It is interesting to compare this result with figures calculated from Pascal's additivity formula

$$\chi_M = \sum \alpha \chi_A + \lambda, \dots\dots\dots (7)$$

where χ_M is the molar susceptibility; α the number of atoms of "atomic susceptibility" χ_A ; and λ the "constitutive correction". The atomic susceptibilities and constitutive factors of Pascal (corrected to the mass susceptibility of water, $\chi' = -0.720 \times 10^{-6}$) were taken on the assumptions that the nitrogen atoms were cyclic and that one oxygen atom of each propionic acid group behaved magnetically as an alcohol oxygen. The usual constitutive corrections were applied to the vinyl and carboxyl groups and the carbon atoms of the porphyrin nucleus were corrected for as mononuclear or binuclear depending on their structural position. The value so calculated for the molar susceptibility (-358×10^{-6}) differs significantly from the observed one.



Fig. 2.—The resonance systems of (a) free porphyrin, (b) covalent haemin, and (c) ionic haemin.

Another calculation was based on a combination of values for Pascal's constants and known values for pyrrole and propionic acid. The molar susceptibility of pyrrole (-48.7×10^{-6}) had been reported by Bonino and Manzoni-Asidei (1943) and of propionic acid (-43.46×10^{-6}). These figures were combined with the Pascal values for vinyl and methene groups and the constitutive corrections not already accounted for in the pyrrole and acid groups, to give a summation (-377×10^{-6}) closer to the experimental value; the difference, however, is still significant. This difference could not be due to incomplete purification of the sample, as haemin is the only impurity likely to be present, and being strongly paramagnetic it would have the effect of lowering the diamagnetic susceptibility. The experimental value is more diamagnetic than the additive values so the presence of haemin cannot account for the discrepancy. It is assumed that the porphyrin ring possesses a characteristic constitutive correction factor. As the iron atoms of haemin compounds can possess ionic or covalent bonds, it is not possible to estimate a specific value for the diamagnetic correction, suitable for all derivatives. The resonance systems will differ (Fig. 2) and so affect the magnetic properties of the atoms involved. This fact does not invalidate use of a correction factor based on pure porphyrin, as the magnetic property of the resonance system of the latter would appear to be intermediate in value to those of the systems shown for ionic and covalent iron-porphyrins.

V. IONIC AND COVALENT IRON BONDS

Iron, as a typical member of the transition group of elements, possesses five unpaired electrons when the ferric form is linked by essentially ionic bonds of sp^3 type (Pauling 1932). In the presence of substances able to form covalent bonds the magnetic moment should decrease, as there must be pairing of electrons. Theoretically, ferric iron possesses either three or one unpaired electrons in its covalent states. In an extension of Pauling's work, Hultgren (1932) has calculated on the reasonable assumption of cylindrical symmetry of bond functions that no more than six equivalent bonds may be formed. The strongest bonds for six covalent functions are directed to the corners of either an octahedron or an equilateral prism with square sides. Four equivalent bonds are most stable when directed to the corners of a square, if only one d eigenfunction is available. The magnetic moments of ferric compounds should, therefore, assume one of the values 5.92, 3.87, or 1.73 Bohr magnetons. There can be no continuous transition of bond type from ionic to covalent, on account of the discrete pairing of electrons. When μ values differ from those predicted, workers have often made one of the following assumptions:

- (i) The orbital moment is not completely quenched. As the normal state of ferric iron is ${}^6S_{5/2}$, no such contribution would be expected, although it has been suggested for many d^2sp^3 type of complexes with μ values of about 2.4.
- (ii) Iron-iron interaction occurs. It has not been established conclusively, although there is strong evidence in the case of iron carbonyl, that such an effect occurs with non-ferromagnetic compounds. It is possible with closely-packed paramagnetic atoms.
- (iii) There is a mixture of components possessing different bond types.

VI. IRON HYDROXIDES AND FERRITIN

The close structural relationship between the haematin, and between the simple oxides and hydroxides of iron, and ferritin, leads us to consider the latter components before dealing with the haematin group. Ferric ions, in aqueous medium, will be coordinated with water molecules (Pauling 1948), and in the presence of alkali hydroxyl ions will compete for coordinating positions. It is well known that the magnetic moment of ferric hydroxide, prepared from simple salts in water, varies from 5.8 to 2.5, and depends upon the conditions of precipitation and subsequent treatment. The ferric minerals goethite, lepidocrocite, and limonite $(\text{FeO.OH})_n$, and haematite (Fe_2O_3) when measured as powders give μ values between 2.7 and 2.4. These variations reflect the ability of iron and oxygen to combine either by ionic or covalent linkages. In most cases it must be assumed that specimens contain mixtures of components each possessing ionic, intermediate covalent, or completely covalent types of bonding. The formation of covalent bonds is attributed to the removal of the elements of water from the hydroxyl groups of adjacent iron atoms, with the consequent production of iron-oxygen-iron bonds. It should be remembered that the most stable covalent configuration for oxygen requires

two bonds directed at about right angles to one another and this has been confirmed in, for example, the crystalline minerals mentioned above.

Michaelis, Coryell, and Granick (1943) found that ferritin, the iron-storage protein, in all preparations studied gave a constant value, $\mu = 3.81$; the ferric hydroxide separated from ferritin gave $\mu = 3.77$. The constant μ values of ferritin iron, almost identical with that expected for covalent dsp^2 bonding, is in distinct contrast to the variations shown by other preparations. These authors stated, "the only objection against this interpretation is that the agreement is almost too good". Perhaps the reason for this comment was based on the fact that this type of bonding has not been conclusively demonstrated among iron compounds, otherwise it would appear odd to deny the case as reasonable evidence in support of the theoretical prediction. In crystalline ferritin, the micelles of ferric hydroxide are assumed to occupy the spaces between the apoferritin molecules, as both ferritin and apoferritin crystallize in identical form and show no significant difference in unit-cell dimensions (Granick 1946). If the iron in ferritin does possess dsp^2 bonds, it would suggest (Pauling 1931) that four stable covalent bonds were directed at right angles in the one plane. The covalent bonding could occur between iron and protein, or iron, oxygen, and iron. Whatever the reason, it does appear that apoferritin exerts a specific influence on the formation of the ferric hydroxide micelles. The constancy of magnetic moment of these micelles even after removal from the protein suggests that the covalent links are between iron-oxygen-iron, because if separation from apoferritin involved the breaking of covalent bonds, the micelle would be expected to show just as wide variations in μ values as those observed with simple ferric hydroxide preparations.

VII. SOLID HAEMINS

The results of Table 2 show that most of the stable iron derivatives possessed magnetic moments very close to the theoretical value calculated by Pauling (1932) for ferric compounds with ionic bonding, which assumes that the orbital moment of the five unpaired electrons is quenched completely. The variations from the theoretical value, although significant, are not sufficient to indicate any type of bonding other than ionic. Coryell, Pauling, and Dodson (1937) showed that methaemoglobin possessed a moment of 5.80 which agrees closely with our mean value for protohaemin chloride of 5.79. They suggested that magnetic interaction might explain the slight lowering of moment from 5.92 expected by theory.

The structural differences between these iron porphyrins might be summarized as

- (i) The anion neutralizing the residual positive charge on the ferric ion, viz. chloride, bromide, formate, acetate, propionate, or hydroxide (see later).
- (ii) The peripheral substituents of the porphyrin ring. In protohaemin the eight substituents are 1,3,5,8 tetramethyl; 2,4 divinyl; 6,7 dipropionic acid. The changes found to have no effect on the bond type are the

esterification of the propionic acid groups, the saturation of the vinyl groups with formation of mesohaemin, and replacement of vinyl groups by hydrogen with formation of deuthaemin.

TABLE 2
MAGNETIC MOMENTS OF SOLID HAEMINS

Derivative	Per cent. Fe (Mean)	Mass Suscept- ibility $\chi \times 10^6$	Dia- magnetic Correction $\chi_M \times 10^6$	Temper- ature (°C.)	Magnetic Moment (μ)	Mean (μ)
1. Protohaemin chloride	8.53	21.56	-416	289.5	5.82	5.79
		21.29		289.5	5.79	
		21.43		287.0	5.78	
		21.45		286.0	5.77	
		21.45		286.0	5.77	
2. Protohaemin chloride dimethyl ester	8.18	20.81	-440	285.2	5.80	5.76
		20.16		286.0	5.72	
3. Protohaemin bromide	7.75	20.77	-429	286.2	5.96	5.91
		19.77		286.9	5.83	
		20.44		287.4	5.93	
4. Protohaemin formate	8.30	22.16	-411	284.0	5.92	5.91
		21.75		286.9	5.90	
5. Protohaemin acetate	8.22	20.59	-423	286.2	5.77	5.82
		21.06		285.6	5.84	
		21.34		283.5	5.84	
6. Protohaemin pro- pionate	8.91	20.69	-435	285.1	5.55	5.48
		19.58		285.1	5.41	
7. Mesohaemin chloride	8.50	22.33	-439	287.9	5.92	5.92
8. Deuthaemin chloride	8.68	21.29	-394	285.3	5.69	5.67
		20.94		285.5	5.65	
9. Azahaemin chloride	8.59	21.29	-415	285.0	5.72	5.74
		21.31		288.0	5.75	
		21.28		287.0	5.74	

(iii) The replacement of one methene bridge by nitrogen with formation of mono-azahaemin. From the results obtained on the substitution of vinyl groups by hydrogen or methyl groups and esterification of propionic acid groups, it must be concluded that peripheral substituents exert no major influence in determining the nature of iron bonds.

A number of points arise from these findings. The nature of the anion can vary a great deal without disturbing the electronic configuration of the metal sufficiently to induce covalent bonding. This fact is not unexpected and it adds supplementary evidence to work done on other ferric compounds. For examples, the inhibition of catalase activity by anions has been explained by Theorell (1946) as a competitive effect between hydroxyl ions (essential for activity) and other anions. Theorell (1943) has also shown that in the case of the active form of peroxidase I the peroxide ion is linked ionically to iron. The inactive form, peroxidase III, and possibly also methaemoglobin produce peroxide complexes with essentially covalent bonds.

VIII. WATER-INSOLUBLE HAEMATINS

The rather close agreement in μ values for the derivatives listed in Table 2 contrasts strikingly with the values of haematin preparations shown in Tables 3 and 4. That is, derivatives with widely differing general constitutions and haematin-type compounds, *when prepared in non-aqueous liquids*, all possess ionically bound iron. In an aqueous environment, however, even when coordinating bases such as pyridine are absent the magnetic moment of haematin-iron can vary between 5.08 and 2.38. The effect of alkali in lowering the magnetic susceptibility of haematin in aqueous solution was first reported by Rawlinson (1940).

TABLE 3
MAGNETIC MOMENTS OF SOLID, WATER-INSOLUBLE HAEMATINS

Sample	Preparation	μ
10	From haemin acetate by hydrolysis	5.81
11	From pyridine haemin by drying etc. ..	5.84
12	Haematin half-anhydride	5.99
13	Haematin anhydride	5.81
14	Barium haematin	5.25
15	From alkaline haematin by precipitation with acid	5.08

As other iron compounds, e.g. ferric hydroxide, ferritin, and some haem-proteins, also show low moments, and as these compounds function biologically in aqueous media, it is important to explain the chemical nature of such observations.

IX. COMPARISON OF HAEMATINS WITH OTHER IRON COMPOUNDS

In haematin derivatives, the iron is situated in the centre of a planar porphyrin molecule, and this steric restriction must be taken into account when considering coordination with other groups. The crystal lattices of the ionic complexes of compounds 1-9 (Table 2), must consist of layers of molecules, with anions neutralizing the residual positive charges on adjacent iron atoms. Haemin

acetate when left in contact with cold water is converted into haematin (compound 10), without visibly affecting the crystal structure; the product possesses a high magnetic moment, 5.81, so it must be concluded that the replacement of acetate does not change the nature of the iron bonds. It is not known whether the acetate is replaced by neutral water molecules or by hydroxyl ions, and it would be of interest to study the mechanism by testing the residual charge on the crystals, or by measuring the pH or conductivity changes in the water used. In contrast, it is not possible to displace chloride from haemin chloride even in boiling water (Richter 1938), and this has been taken as evidence of the inability of water to penetrate the lattice layers (Lemberg and Legge 1949). The complete explanation must take other more fundamental factors, such as ionic potential, into account.

The haematin (compound 11) prepared by heating crystalline pyridine haematin, also possesses a high magnetic moment, indicating the presence of ionic bonds, probably for the same reasons as discussed for haemin acetate. In the presence of water, pyridine is able to form covalent complexes with haematin (Rawlinson 1940) so there is no point in attempting to replace it in the same way as for acetate.

The magnetic moments of haematin 15-19, prepared from aqueous alkali, offer several interesting points for comment. The preparation (16) which was obtained from dilute KOH solution by precipitation on the addition of acid, possessed a magnetic moment (5.08) sufficiently low to indicate clearly that

TABLE 4
MAGNETIC MOMENTS OF WATER-SOLUBLE HAEMATINS PREPARED UNDER
DIFFERENT CONDITIONS

Sample	Preparation	Magnetic Moments (μ)	
		Solid	Aqueous Solution
16	Precipitated in 9.5M aqueous KOH	2.40	—
17	Dry precipitate from 5M KOH	2.92	3.55
18	Dry precipitate from 7.5M KOH	2.38	2.85
19	Dry precipitate from methanol-KOH	5.70	3.42

more than one component was present, and that one of them contained covalent iron bonds. In haematin prepared under these conditions, it is likely that water molecules rather than hydroxyl ions are coordinated, and other evidence based on the dissociation constants of groups supports this suggestion (Lemberg and Legge 1949, p. 170).

The main conclusions to be drawn from the results listed in Table 4 are that iron forms (a) an intermediate type of covalent bonding under mildly alkaline conditions, and (b) completely covalent bonding in concentrated alkali. The μ

values of samples 17 and 19 when dissolved in water are in agreement with those obtained earlier (Rawlinson 1940). The precipitates, 16 and 18, have moments similar to well-established d^2sp^3 covalent iron complexes; the difference between the experimental values of 2.4 and the theoretical moment of 1.73, is attributed either to some contribution from orbital moment or to magnetic interaction between iron atoms in close proximity. As the potassium salt of haematin precipitated from 9.5M KOH solution possessed the same moment, while still suspended in the mother liquor, as the specimen obtained from 7.5M KOH after washing with ethanol and drying, it is obvious that the covalent bonds were not changed by treatment after precipitation. The μ value of compound 18

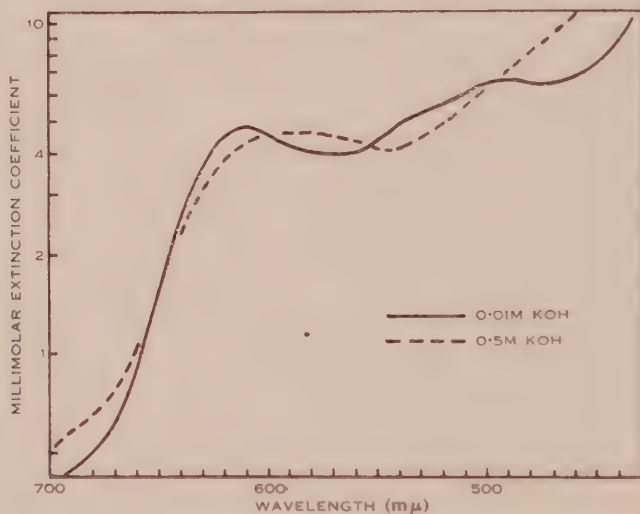


Fig. 3.—The two types of absorption spectra of haematin in aqueous alkali.

is slightly greater than 2.4, and probably means that conversion was incomplete. The value, 2.85 for sample 19 when dissolved in water, is taken to indicate that insufficient time was allowed for regeneration of the intermediate type of covalent bonding.

These conclusions are supported by spectrophotometric evidence: between alkali concentrations of 0.01 and 0.3M the absorption spectrum of haematin does not change significantly, but above about 0.4M another type of spectrum appears (see Fig. 3). This spectrum also remains unchanged up to at least 1M, after which interference produced by precipitation makes measurements unreliable. The presence of isosbestic points at 650, 601, 553, and 497 mμ offers unequivocal proof for the existence of only two spectroscopically distinct components. The colour of solutions also changes with changes in spectral type and magnetic moment: at low alkali levels the solutions are brownish red, and at higher levels reddish brown. The reversion of the high alkali component to the other component does not occur spontaneously by dilution. It is probably associated with a hydration process.

Hartree (1946) suggested that polymerization was responsible for the low μ values of haematin in alkali. This suggestion implies that there would be a continual transition in the absorption spectrum of haematin with increasing concentration of alkali, caused by change in particle size. It is certain that polymerization does occur but this fact is the result of, rather than the cause of, covalent iron bond formation. Hogness *et al.* (1937) and Davies (1940) concluded from spectrophotometric data that haematin in alkali was dimeric but Barron (1937) was unable to confirm this from oxidation-reduction studies. From our present evidence it is impossible to assume more than the existence of polydisperse particles. In order to give a useful basis for discussion, diagrams are presented in Figure 4. The dimensions have been drawn to scale so as to represent as accurately as possible the ionic, covalent, and van der Waals radii of atoms. This information makes it less difficult to study the steric and geometric relationships of the various structures. In Figure 4 (a) it can be seen how iron binds ionically to pyrrole nitrogens and chloride, to give parallel packing of planar haemin molecules in crystals. In solutions below pH 5 the propionic acid groups would be undissociated and the unit positive charges on the iron atoms would be neutralized by anions (Fig. 4 (b)), and the sixth coordination position would probably be occupied by a water molecule. The propionic acid groups would be fully dissociated at neutral reaction, and in the absence of anions two water molecules should be coordinated (Fig. 4 (c)). As the solutions become more alkaline then hydroxyl ions will bind to iron, although in buffered solutions other anions will compete to some extent; Figure 4 (d) illustrates the monomeric form of haematin, and Figures 4 (e) and 4 (f) represent two dimeric forms.

The flat haematin molecules in aqueous environment would, with increasing alkali concentration, be subjected to stronger ionic forces and as a result the iron must be induced to form covalent bonds (Figs. 4 (d)-(i)). The four covalent bonds, presumably of dsp^2 type, must be directed in one plane at 90 or 180° to one another, and to satisfy this condition in such a sterically restricted molecule there must be either (i) a sharing of electrons between iron and the nitrogens of the porphyrin nucleus, or (ii) the formation of covalent bonds with oxygens both above and below the plane of the nucleus. Each of these possibilities is reminiscent of those discussed for ferritin iron. The first possibility would not be expected to encourage polymer formation, but it would suggest a definite change in the nature of the absorption spectrum. The other possibility would result in polymerization without producing such great spectral changes, as the bonds from iron to porphyrin would still remain essentially ionic. The magnetic moment, $\mu=2.4$, of haematin in concentrated alkali indicates the completion of covalent bond formation to iron. The explanation of this structure is very difficult to make in view of several restrictions. The rigid, planar nucleus permits only parallel packing of molecules in polymers, and the bonds from iron to pyrrole nitrogens must be at 90 and 180° within the plane. If each oxygen is linked covalently between iron atoms, then it would not be expected that its bond angles were directed at 180° (see Figs. 4 (g) and 4 (h)). The attainment of ideal octahedral bond directions would therefore be denied for iron atoms. If

two oxygens were assumed to be bound on each side of iron, then another octahedral arrangement is possible by assuming that only two haematin nitrogens were involved in covalent bond formation. One or more oxygens could be bound covalently between each iron atom, as shown in Figures 4 (e)-(i). The latter types of structure (Figs. 4 (h) and 4 (i)) would tend to be formed in concentrated alkali by the loss of protons from bound to free hydroxyl ions with the formation of water. Figure 4 (i) is included to show the closest packing possible

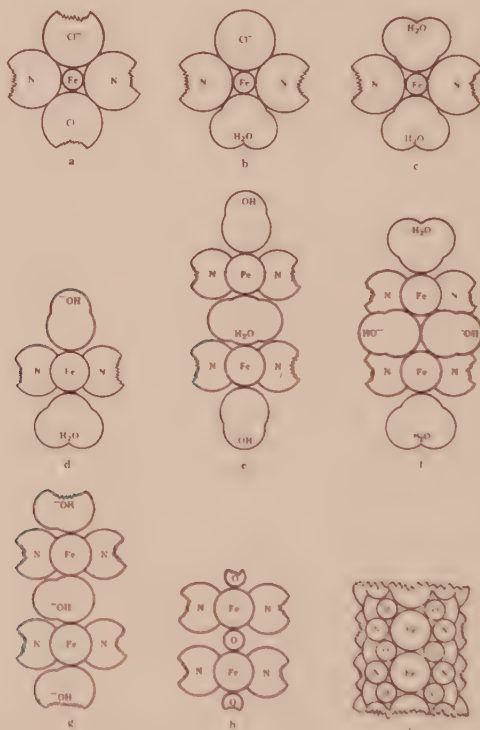


Fig. 4.—The structural nature of haem compounds.

when haematin polymerizes by the formation of iron-oxygen-iron bonds, while at the same time maintaining stable covalent oxygen bond angles. The other restriction concerns the maximum number of covalent bonds to each iron atom. Hultgren (1932) concluded as mentioned above, that not more than six stable bonds could be formed. If this is the case, and if a distorted octahedral bond structure is unstable, then the only other stable configuration would appear to be based on a distribution of iron bonds to the corners of a trigonal prism. Such a structure would be possible by the combination of iron with three equally spaced oxygen atoms, both above and below the plane of each haematin molecule but it would not allow the formation of covalent bonds between iron and pyrrole nitrogens. The estimation of distances between iron-iron, and iron-oxygen would be invaluable in defining the nature of covalent bonds in such haematin complexes.

X. HAEMATIN ANHYDRIDES

The work of Hamsik (1936) has drawn attention to formation of crystalline products containing less oxygen than the haematin from which they were derived. Haematin anhydride and half-anhydride were claimed to be formed by the elimination of one molecule of water from one or two molecules of haematin, respectively. The physical characteristics of these preparations certainly differ from freshly prepared haematin, and from one another. The various possibilities of anhydride formation might be summarized as

- (i) The elimination of water between propionic acid groups, but if this were so then haemin chloride should form similar products.
- (ii) The elimination of water between propionic and hydroxyl groups of the same or adjacent molecules; Hamsik (1923-4) considered that haematin anhydride is formed in this way.
- (iii) The elimination of water between hydroxyls of adjacent haematin molecules. The participation of iron-bound oxygen atoms in anhydride formation seems likely. The ionic nature of the iron bonds in these derivatives is not of much interest unless anhydride formation is proved in aqueous medium, because haematin prepared in organic solvents also possesses ionic iron bonds.

It is obvious that a full explanation of structures in iron-porphyrin compounds must await the detailed development of general theories of chemical bonding. The stimulation and practical contributions of Pauling (1948) are well known, and it is to be hoped that findings such as those presented here, on changes in simple derivatives of natural substances in aqueous medium, might encourage further work to assist in the elucidation of the biological functions of haem compounds.

XI. ACKNOWLEDGMENT

The Australian National Health and Medical Research Council was responsible for the full financial support of this work.

XII. REFERENCES

- BARRON, E. S. G. (1937).—*J. Biol. Chem.* **121** : 285.
BONINO, G. B., and MANZONI-ASIDEI, R. (1943).—*Ber. dtsch. chem. Ges.* **76** : 553.
CAMBI, L., and SZEGO, L. (1934).—*R.C. Ist. Lombardo* **67** : 275.
CORYELL, C. D., PAULING, L., and DODSON, R. W. (1937).—*J. Amer. Chem. Soc.* **59** : 633.
DAVIES, T. H. (1940).—*J. Biol. Chem.* **135** : 597.
DRABKIN, D. L. (1941).—*J. Biol. Chem.* **140** : 387.
FISCHER, H., and HUMMEL, G. (1929).—*Hoppe-Seyl. Z.* **181** : 127.
FISCHER, H., and PUTZER, B. (1926).—*Hoppe-Seyl. Z.* **154** : 17.
FISCHER, H., TREIBS, A., and ZEHLE, K. (1930).—*Hoppe-Seyl. Z.* **193** : 160.
GRANICK, S. (1946).—*Chem. Rev.* **38** : 379.
GRAY, F., CLOW, A., and CRUICKSHANK, J. (1936).—*J. Sci. Instrum.* **13** : 13.
HAMSIK, A. (1923-4).—*Publ. Fac. Méd. Brno* **2** : 1.
HAMSIK, A. (1924).—*Hoppe-Seyl. Z.* **133** : 173.
HAMSIK, A. (1929a).—*Hoppe-Seyl. Z.* **183** : 103.
HAMSIK, A. (1929b).—*Hoppe-Seyl. Z.* **182** : 117.

- HAMSIK, A. (1930).—*Hoppe-Seyl. Z.* **190** : 211.
- HAMSIK, A. (1936).—*Hoppe-Seyl. Z.* **241** : 156.
- HARTREE, E. F. (1946).—*Ann. Rep. Chem. Soc. London* **43** : 287.
- HAUROWITZ, F., and KITTEL, H. (1933).—*Ber. dtsh. chem. Ges.* **66B** : 1046.
- HOGNESS, T. R., ZCHIELE, F. P., SIDWELL, A. E., and BARRON, E. S. G. (1937).—*J. Biol. Chem.* **118** : 1.
- HULTGREN, R. (1932).—*Phys. Rev.* **40** : 891.
- KÜSTER, W. (1927).—*Hoppe-Seyl. Z.* **163** : 272.
- LEMBERG, R. (1943).—*Aust. J. Exp. Biol. Med. Sci.* **21** : 239.
- LEMBERG, R., and LEGGE, J. W. (1949).—"Haematin Compounds and Bile Pigments." 1st Ed. p. 168. (Interscience Publishers Inc. : New York.)
- MICHAELIS, L., CORYELL, C. D., and GRANICK, S. (1943).—*J. Biol. Chem.* **148** : 463.
- PAULING, L. (1931).—*J. Amer. Chem. Soc.* **53** : 1367.
- PAULING, L. (1932).—*J. Amer. Chem. Soc.* **54** : 988.
- PAULING, L. (1948).—*J. Chem. Soc.* **1948** : 1461.
- PAULING, L., and CORYELL, C. D. (1936).—*Proc. Nat. Acad. Sci.* **22** : 159.
- RAMSAY, W. N. M. (1944).—*Biochem. J.* **38** : 467.
- RAWLINSON, W. A. (1940).—*Aust. J. Exp. Biol. Med. Sci.* **18** : 185.
- RICHTER, A. F. (1938).—*Hoppe-Seyl. Z.* **253** : 193.
- SELWOOD, P. W. (1943).—"Magnetochemistry." 1st Ed. (Interscience Publishers Inc. : New York.)
- THEORELL, H. (1943).—*Ark. Kemi Min. Geol.* **16A** (14) : 1.
- THEORELL, H. (1946).—*Arch. Biochem.* **10** : 321.

THE SURFACE CHEMISTRY OF PROTEINS

III. THE RIGIDITY OF ADSORBED FILMS AT AN AIR-WATER INTERFACE

By C. W. N. CUMPER* and A. E. ALEXANDER*

[*Manuscript received August 29, 1951*]

Summary

The results presented in Part II of this series on the adsorption of the proteins, β -globulin, pepsin, and insulin at a white oil, 5 per cent. oleyl alcohol-water interface have been extended to the petrol ether-water and air-water interfaces by studying the surface rigidity of the films formed.

The previous picture of the process of adsorption has been generally confirmed but there are certain differences between the oil-water and air-water interfaces which are discussed.

I. INTRODUCTION

The adsorption and denaturation phenomena of proteins at an air-water (A/W) interface cannot be examined by the surface viscosity technique used at the oil-water (O/W) interface (Cumper and Alexander 1950), since the films are much thicker and consequently too viscous. They can, however, be studied by measurement of surface rigidity. The present paper deals with the results so obtained; in a general way they support the conclusions drawn from the previous work although some differences were observed which will be discussed.

II. EXPERIMENTAL

Direct measurement of the shear rigidity of the film by a static method was first attempted. A mica disk was suspended from a torsion wire so that it lay in the water surface at which the protein was adsorbing. A torque was applied by means of a torsion head and the movement of the disk followed with a lamp and scale. When the torque was small, the disk did not move; when it was increased, the disk rotated until ultimately it was under no shearing stress. This method was therefore abandoned in favour of that used by Mouquin and Rideal (1927).

In their method a viscous drag is exerted upon a circular film firmly anchored at its edges by a disk rotating beneath the film, and the displacement of the film measured directly by observing the movement of fine talcum particles on its surface. To eliminate superfluous mechanism and contamination in the water a perspex disk (radius 5 cm.) was driven by a spindle, coated with paraffin wax, passing through the surface. This spindle was loosely encased in a glass tube so that it was not in direct contact with the film. The deflection of the talcum powder was measured by a grating in the microscope eyepiece. In this way

* School of Applied Chemistry, N.S.W. University of Technology, Sydney.

displacements between 0.006 and 0.06 cm. could be measured to about 0.0005 cm. and the applied stress was adjusted so that the deflection was always

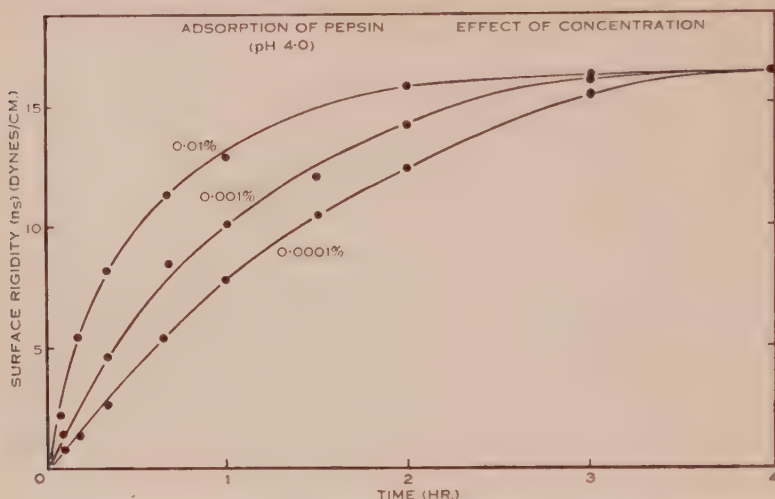


Fig. 1.—The surface rigidity of pepsin adsorbed at an A/W interface (pH 4.0) as a function of time at three concentrations.

within this range. With this method it was easy to see if the elastic limit of the film had been exceeded, as then the talcum particles did not return to their original position when the stress was removed.

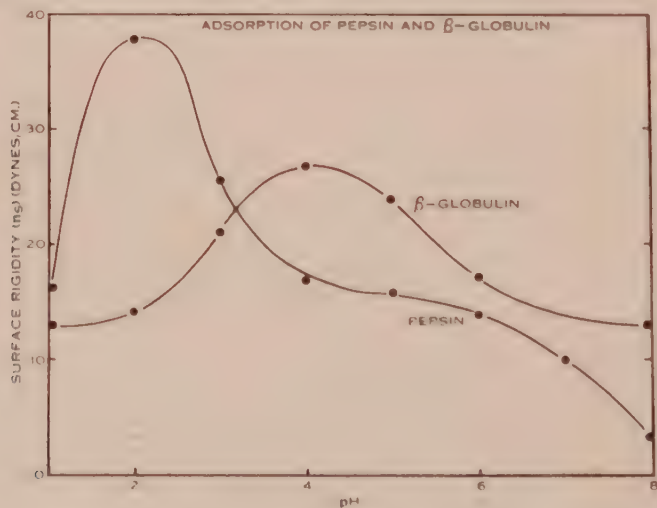


Fig. 2.—The surface rigidity of pepain and bovine β -globulin adsorbed for 4 hours at an A/W interface as a function of pH.

The surface rigidity (n_s) is calculated from the equation

$$n_s = -\frac{\omega\eta}{8h\delta}(r^3 - a^2r),$$

where ω is the angular velocity of the disk of radius a rotating at a depth h beneath the film, and δ is the displacement at a distance r from the centre of the film. In Table 1 observed displacements (δ) and the calculated rigidity are recorded for various values of r and h . The accuracy was about 5 per cent. except for the largest values of r where the displacements were too great. This was due to the radius of the disk being 1 cm. less than that of the film (6 cm.) which was necessary to minimize stray currents along the sides of the dish.

TABLE 1
OBSERVED DISPLACEMENTS AND CALCULATED RIGIDITY
 $2\pi/\omega = 7.60$ sec. $\eta = 0.009$ poise $a = 5.0$ cm.

r (cm.)	h (cm.)	δ (cm.)	n_s (dynes/cm.)
5.00	0.25	0.0064	—
4.42	0.30	0.0106	7.0 ₁
	0.25	0.0128	6.9 ₇
	0.15	0.0190	7.8 ₃
3.84	0.70	0.0064	8.1 ₂
	0.25	0.0160	9.1 ₅
3.26	0.90	0.0064	7.5 ₄
	0.25	0.0212	8.1 ₈
	0.18	0.0276	8.7 ₀
2.68	0.82	0.0060	9.0 ₀
	0.25	0.0221	7.9 ₈
2.44	0.53	0.0094	8.6 ₅
	0.25	0.0212	7.8 ₈
	0.17	0.0320	7.0 ₇
Mean			8.0 ₈ \pm 0.5

The same proteins were used as in Part II of this series, viz. bovine β -globulin, insulin, and pepsin (all from Armour and Co. Pty. Ltd.) and, unless stated otherwise, the ionic strength of the solution was 0.03 , temperature 25 ± 0.1 °C., and protein concentration 0.01 per cent.

III. RESULTS

(a) Effect of Protein Concentration

Figure 1 shows the surface rigidity plotted as a function of the time of adsorption for three concentrations of pepsin in an acetate buffer of pH 4.0. Concentration affects the rate at which the film builds up but not the final value of the surface rigidity. This was not the case at the O/W interface (Cumper and Alexander 1950); a possible explanation will be given later.

(b) *Effect of pH*

The surface rigidities of β -globulin and pepsin films adsorbed at various pH values are shown in Figure 2. The proteins were adsorbed for 5 hours but there

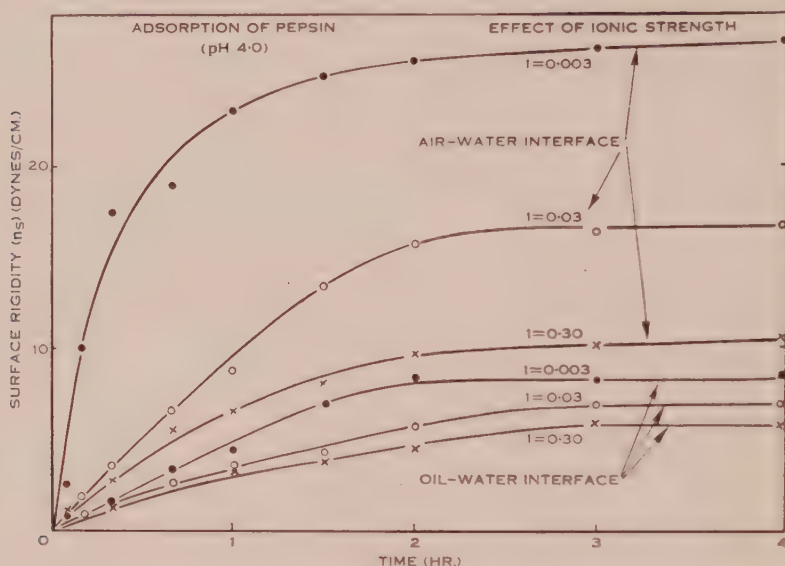


Fig. 3.—The surface rigidity of pepsin adsorbed at the A/W and petrol ether-water interfaces (pH 4.0) as a function of time at three ionic strengths.

was little change in the rigidity after only 3 hours. These results are similar to those obtained for the surface viscosity of these films (Cumper and Alexander 1950) but the maxima in the curves are not at the same pH (Table 2).

TABLE 2
COMPARISON OF pH MAXIMA

Protein	pH of Maximum Surface Viscosity at O/W Interface	pH of Maximum Rigidity at A/W Interface	Isoelectric Point
β -Globulin	3.2	4.0	5.0
Insulin	5.2	—	5.0
Pepsin	3.2	2.0	?
(see Cumper and Alexander 1950)			

With insulin the film flowed at all pH values and the rigidity could not be measured, although it was noticeably more viscous near its isoelectric point.

(c) Effect of Ionic Strength

The effect of ionic strength upon the surface rigidity of pepsin films at pH 4.0, adsorbed at the petrol ether-water and at the A/W interfaces, was studied (Fig. 3). The presence of the oil lowered the interfacial tension from 72 to 48 dynes/cm. and reduced the surface rigidity by a factor of about two. However, the general shape of the curves remained the same being similar to those obtained for the surface viscosity at the white oil, 5 per cent. oleyl alcohol-water interface (Cumper and Alexander 1950).

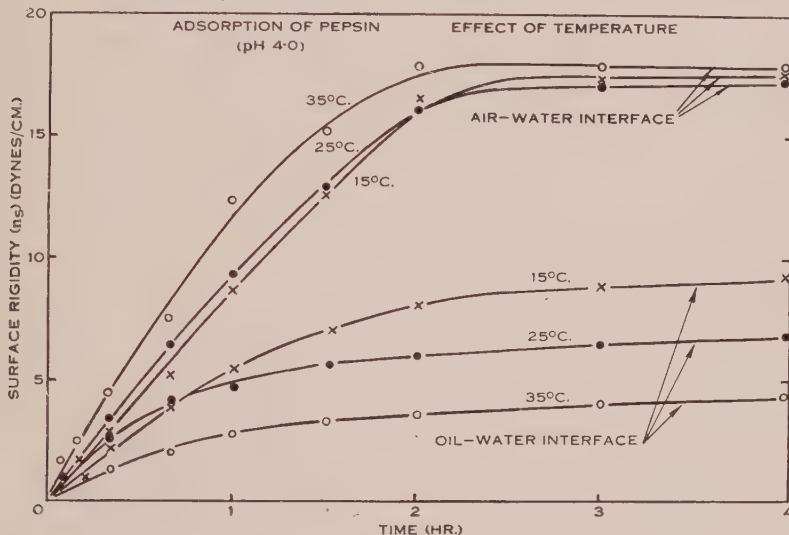


Fig. 4.—The surface rigidity of pepsin adsorbed at the A/W and petrol ether-water interfaces (pH 4.0) as a function of time at three temperatures.

(d) Effect of Temperature

The corresponding results for the effect of temperature on the adsorption are given in Figure 4. At the O/W interfaces the surface rigidity and viscosity are affected similarly by a change of temperature. But at the A/W interface varying the temperature from 15 to 35 °C. had little effect on the surface rigidity. However, with a pepsin film formed by adsorption for 5 hours at pH 7.0 and therefore in a stationary state, an increase in temperature produced a marked decrease in the surface rigidity (see Fig. 5). This effect was reversible.

IV. DISCUSSION

The principal experimental results which require elucidation can be summarized as follows:

- The rigidity becomes constant after about 3 hours, at a value which is independent of the protein concentration.
- The rigidity-pH curve goes through a maximum at a pH near to, but not identical with, either the isoelectric point or the pH of maximum viscosity at the O/W interface.

- (c) Increasing the ionic strength decreases the equilibrium value of the rigidity, the effect being more pronounced at the A/W than at the O/W interface.
- (d) Under otherwise identical conditions, replacing the air by oil markedly reduces the equilibrium value of the rigidity.
- (e) Increasing the temperature increases the equilibrium rigidity at the O/W interface, but has little effect at the A/W interface.

The above results indicate that, in a general way, the surface rigidity behaviour of adsorbed protein films runs parallel to the surface viscosity of

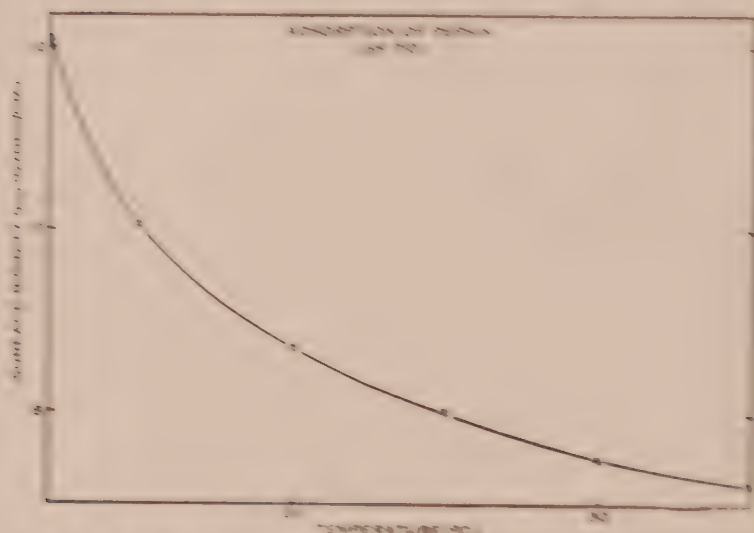


Fig. 1.—The effect of temperature on the surface rigidity of a protein film adsorbed at an A/W interface for 5 hours (pH 7.0).

the white oil, 5 per cent. vinyl alcohol-water interface (Jumper and Alexander 1959). The general picture of adsorption of the native protein molecule, followed by swelling (surface deterioration), and subsequently by coagulation, as described previously by Jumper and Alexander (1959) would therefore seem to be applicable to the A/W interface also. However, the conditions in the two cases are not identical for the following reasons:

- (i) The surface energy of the A/W interface is much greater than that of the white oil, 5 per cent. vinyl alcohol-water interface and consequently the adsorbed films tend to be thicker.
- (ii) The surface viscosity technique involved considerable disturbance of the film and this would certainly promote coagulation of the surface denatured chains into small discrete aggregates which can be readily displaced from the interface (coagulation). Such disturbance is effectively absent in the surface rigidity method.

- (iii) At an O/W interface the hydrophobic coagulum can be removed from the denatured film much more readily; also the polypeptide chains in the surface denatured form will be much more mobile owing to the marked reduction in the van der Waals adhesive forces.

After a detailed consideration of various possible explanations of phenomena (a) to (c) above, we have concluded that, under the undisturbed conditions used for the surface rigidity measurements, only surface denaturation is occurring and the formation of surface coagulated material as defined above cannot occur. At a given interface the equilibrium film thickness is assumed to be essentially independent of pH, salt concentration, protein concentration, and temperature, being determined almost entirely by the magnitude of the initial interfacial energy (air-water 72 erg cm.²; petrol ether-water 48 erg cm.²; white oil, 5 per cent. oleal alcohol-water 15.3 erg cm.²). Such a view provides a ready explanation for certain phenomena termed "stability" and "permanence" in protein stabilized emulsions (Aberne and Reilly 1944).

The adsorption process is pictured as follows. The adsorbed native molecules undergo surface denaturation in the extremely thin region (c. 5 Å thick) which constitutes the discontinuity between the bulk aqueous phase and the surface denatured film, and as this proceeds the interface gradually becomes filled with the more hydrophilic groups of the polypeptide chains, of which the ionized side-chains would be the most important. Should the spreading pressure of the native protein be insufficient to force the most hydrophilic of these groups out of the interface then adsorption must eventually come to a stop. The high viscosity of the film (~ 20 surface poises, which corresponds to a bulk value of c. 10^2 poises) must slow down the adsorption process considerably, particularly in its later stages. For the denatured chains to take up the positions of optimum association necessary for coagulation, some relative movement will in general be required to bring appropriate groups into proximity (e.g. $>CO$ and $HN-$, COO^- , and H_3N^+ , etc.). Consequently, at a quiescent interface, as soon as adsorption ceases, the extremely high viscosity of the film will effectively prevent any further coagulation. On the other hand, at a stirred interface, as in the viscosity experiments, the mechanical agitation will enable coagulation to proceed.

The above picture fits in well with the observed rigidity-time behaviour at various protein concentrations (Fig. 1) and also with the reversibility of the rigidity-temperature curve shown in Figure 5.

Considering the adsorption of the native protein in more detail, it is important to find out whether the whole of the molecule, or only a small part of it, must enter the surface before uncoiling (surface denaturation) can occur. The former would require making a "hole" in the surface film of approximately 1000 Å^2 , the latter one of say 40 Å^2 (sufficient to allow free passage of a single polypeptide chain). With a surface pressure π of 30 dynes cm., the corresponding surface-free energies would be 43 kcal. g. mole and 1.7 kcal. g. mole respectively, corresponding to values of the exponential factor $e^{-\pi a/RT}$ of 4.7×10^{-10} and 6×10^{-3} . These figures would seem to preclude the one-stage adsorption of the

whole molecule and make it likely that all steps in the surface denaturation process involve only a small portion of the molecule.

The general form and in particular the maximum of the rigidity-pH curve (Fig. 2) is explained as follows. There will be a certain pH value at which the electrostatic attractive forces between the denatured chains will be a maximum corresponding to the maximum "energy density" of salt bridges (i.e. taking into account strength as well as concentration of each type of salt bridge). This pH, the maximum in Figure 2, would be expected to be in the region of the isoelectric point but its precise position would be influenced by steric and ionization factors as discussed in the case of the bulk rigidity of gelatin gels (Cumper and Alexander 1952). The discrepancy between the pH maxima in the rigidity and viscosity methods can be ascribed to the latter involving the coagulation process, for which the pH dependence might well be different.

The reduction in rigidity on addition of salts (Fig. 3) is ascribed to the reduced electrostatic attractions between the salt bridges. This explanation is supported by the fact that the final rigidities at the O/W interface approximately fit the equation $n_s = (n_s)_0 (1 - 0.77\sqrt{I})$ taking $(n_s)_0$ as 10 dynes/cm., this being the relation expected from the rigidity dependence of concentrated gelatin gels upon salt concentration (Cumper and Alexander 1952). The data for the A/W films do not fit a relationship of this type so well, probably owing to their greater thickness and the much closer packing of the chains. We estimate the protein concentration in the adsorbed films to be about 850 g./l., which would correspond to gelatin gels of very much higher concentrations than those used for comparison (450 g./l.).

The change from an A/W to an O/W interface would be expected to affect the rigidity in two ways: firstly, the films will be thinner owing to the lower interfacial tension and hence lower adsorptive tendency; secondly, the presence of the oil reduces the van der Waals attractive forces between the hydrophobic side-chains which must make some contribution to the overall rigidity.

There remains the variation of rigidity with temperature (Fig. 4). It must first be pointed out that at the O/W interface the rigidity closely parallels the viscosity behaviour (Cumper and Alexander 1950); the apparent anomaly is the comparative insensitivity to temperature of the very rigid A/W films shown at the top of Figure 4. These A/W films will be considerably thicker and have a much closer chain packing than the O/W ones formed under the same conditions (see above), so that it will be more difficult to break cross-linkages than in the case of the thinner and less tightly packed systems. Expressed in another way, in strongly bonded gels the cross bonds are so close that they tend to reinforce one another (i.e. a cooperative effect) whereas in very weak gels each bond can, to a first approximation, be considered separately. Unfortunately there seem to be no data on the temperature variation of extremely concentrated gels to test the validity of this idea.

V. ACKNOWLEDGMENTS

The authors would like to thank Armour & Co. Pty. Ltd. for a gift of the proteins used and the Goldsmiths' Company for the award of a travelling scholarship to one of them (C.W.N.C.).

VI. REFERENCES

- AHERNE, J. C., and REILLY, J. (1944).—*Sci. Proc. R. Dublin Soc.* **23**: 247, 300.
CUMPER, C. W. N., and ALEXANDER, A. E. (1950).—*Trans. Faraday Soc.* **46**: 243.
CUMPER, C. W. N., and ALEXANDER, A. E. (1952).—*Aust. J. Sci. Res. A* **5**: 146, 153.
MOUQUIN, H., and RIDEAL, E. K. (1927).—*Proc. Roy. Soc. A* **114**: 690.

THE PRODUCTION OF INTERMEDIATES IN THE MANUFACTURE OF ATEBRIN

By W. DAVIES,* N. H. OLVER,* and B. W. WILSON*

[Manuscript received July 16, 1951]

Summary

During World War II methods of preparation, suitable for Australian conditions, of the necessary intermediates for the production of the antimalarial drug, atebirin, were investigated. Suitable methods for the preparation of 2,4-dichlorobenzoic acid and *p*-nitroanisole have been developed. 2,4-Dichlorobenzoic acid is prepared from *p*-toluenesulphonyl chloride by chlorination to 2,4-dichlorobenzenesulphonyl chloride (via 2-chloro-*p*-toluenesulphonyl chloride) and hydrolysis of this compound. *p*-Nitroanisole is prepared from *p*-chloronitrobenzene by the action of methanolic potassium, good yields of a pure product being obtained in a shorter time than in methods previously described.

A new synthesis of 5-diethylamino-2-aminopentane was commenced.

I. INTRODUCTION

Early in 1942, when Japan occupied Java and the Allied Nations were thus almost entirely cut off from supplies of quinine, chemists were asked to devise methods for producing, with the chemicals and equipment available in Australia, the necessary intermediates for the preparation of the antimalarial drug atebirin (I). An account of this substance is given by Wiselogle (1946).

Atebrin is prepared by the combination of the aliphatic side-chain, 5-diethylamino-2-aminopentane (II), with 2-methoxy-6,9-dichloroacridine (III), derived by the action of phosphorus pentachloride on 2-methoxy-6-chloroacridone (IV). The latter is obtained by the condensation of 2,4-dichlorobenzoic acid (V) with *p*-anisidine (VI), which is the reduction product of *p*-nitroanisole.

The present paper is an account of investigations for the convenient preparation, under Australian conditions, of the following vital intermediates: (a) 2,4-dichlorobenzoic acid (V), (b) *p*-nitroanisole, and (c) 5-diethylamino-2-aminopentane (II).†

II. METHODS OF PREPARATION

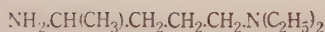
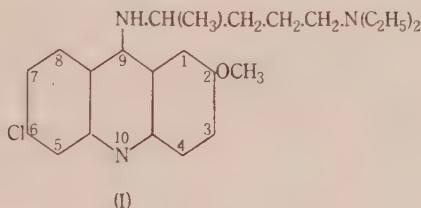
Apart from the chemistry of the processes involved, the methods adopted have an added interest as they fulfilled, as far as possible, the following requirements imposed by the nature of the emergency. The processes should (i) depend on materials available locally, (ii) be rapid, and (iii) use substances without solvents, or else solutions so concentrated that the small apparatus available in universities and colleges would give a large output.

* Organic Chemistry Laboratory, University of Melbourne.

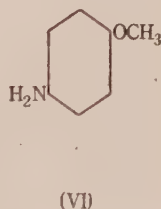
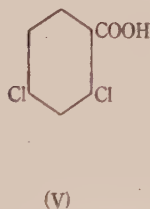
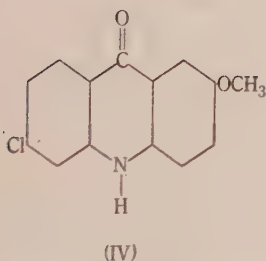
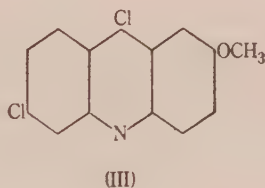
† Further details and references are to be found in the theses for the degree of Master of Science in the University of Melbourne, by N. H. Olver (1942) and B. W. Wilson (1943).

(a) 2,4-Dichlorobenzoic Acid (V)

Of the seven methods for the preparation of 2,4-dichlorobenzoic acid (V) described in the literature, one of the most feasible (and moreover, used overseas in the production of atebtrin) seemed the conversion of aniline, via acetanilide, into 2,4-dichloroaniline, and thence by a Sandmeyer reaction into 2,4-dichlorobenzonitrile which is then hydrolysed to V. Because of the limitations mentioned above, this method was not chosen as the numerous steps take much time and



(II)

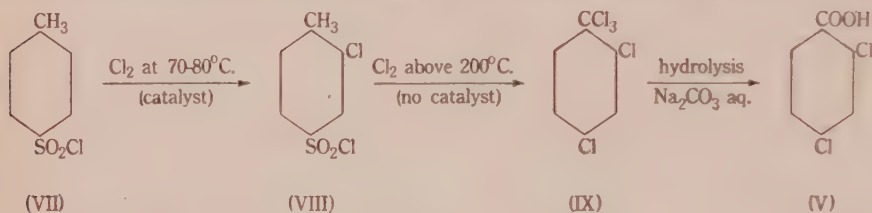


give only a moderate overall yield ; also it necessitates many filtrations involving large volumes of solutions, and consequently large apparatus, which was not available.

Requirements (i), (ii), and (iii) are met by the chlorination of *p*-toluenesulphonyl chloride (VII) to give (via 2-chloro-*p*-toluenesulphonyl chloride (VIII)) 2,4-dichlorobenzotrichloride (IX), which is then hydrolysed to V. The reactions involved are based on various German patents, especially D.R. Pats. 133,000 (1901), 145,908 (1902), and 234,290 (1909), which have been extended by Davies (1921), and Davies and Dick (1932), and D.R. Pat. 639,578 (1935).

In this method the starting product (VII) and the reagents used—chlorine and sodium carbonate—are cheap and readily available, the number of operations is small, and the products are all treated as liquids and distilled until IX is obtained. The catalyst chiefly used is antimony trichloride, which, however, cannot be completely removed from the less volatile VIII by fractionation, and has to be washed out from the organic material, dissolved in a solvent, which

has then to be dried and removed. An examination of a number of catalysts has shown the value of thalious chloride, which is apparently a novel catalyst for the chlorination of aromatic compounds. Its use has the supreme advantage that it is left completely behind when VIII is distilled, and this is then chlorinated at higher temperatures without further interruption and the necessity for the tedious operations described above. The overall yield is excellent, V being obtained from about the same weight of the starting material (VII).



As previously made IX (D.R. Pat. 234,290 (1909)) is described as a liquid, b.p. 155–159 °C./20 mm., but by the present process it is obtained finally in a pure state as a solid, m.p. 50 °C. As impurity, IX is apt to contain a little 2,4-dichlorobenzylidene dichloride. An account of new derivatives which characterize the corresponding aldehyde is given in Section III.

(b) *p*-Nitroanisole

One of the simplest methods for preparing *p*-nitroanisole is by the methylation of *p*-nitrophenol in alkaline solution using dimethyl sulphate. By adding an excess of dimethyl sulphate in successive stages, *p*-nitroanisole is obtained in 82 per cent. yield. However, this method is not really satisfactory because large quantities of dimethyl sulphate are used, and *p*-nitrophenol can be isolated in only poor yields from the mononitration products of phenol.

The availability of locally manufactured *p*-chloronitrobenzene indicated this to be the best starting material. Its conversion into *p*-nitroanisole by means of methanolic sodium methylate has long been known (Holleman and de Mooy 1915), but this is best done by using very dilute solutions, which makes the reaction a very slow one (cf. U.S. Pat. 1,875,916 (1926)). The use of concentrated solutions and a slightly higher temperature brings about the reduction of the *p*-nitroanisole formed, *p*-azoxyanisole being obtained in 64 per cent. yield (Davies and Down 1929).

The use of methanolic potassium or sodium hydroxide instead of sodium methylate is preferable on a manufacturing scale, though here the reaction is complicated by the formation of azoxy-compounds (chiefly *pp'*-dichloroazoxybenzene) as well as by *p*-nitrophenol.

The present authors have verified the formation, in a few minutes, of a 93 per cent. yield of *p*-nitroanisole by the action of methanolic potassium hydroxide on analysed *p*-fluoronitrobenzene as described by Rarick, Brewster, and Dains (1933). This reaction, whilst of little immediate technical value, makes a very informative contrast with the process of Richardson (1926), who obtained a 96 per cent. yield from *p*-chloronitrobenzene and 0.5N methanolic

potash by heating at 60 °C. for 140 hours. This prolonged process using very dilute solutions is not suitable where large-scale apparatus is limited. The patent, U.S. Pat. 1,875,916 (1926), records an improvement, though the time here is 48 hours.

The present investigators have succeeded, using a relatively rapid process and more concentrated solutions, in obtaining an 80 per cent. yield of *p*-nitroanisole in 16 hours. The amount of reduction products is small because the reaction is interrupted before all the *p*-chloronitrobenzene has reacted; this is distilled off and added to the next batch.

In this work it became desirable to find a method of estimating the *p*-nitroanisole and unchanged *p*-chloronitrobenzene present during a series of experiments in which the time and concentrations of reactants were varied. This was done by preferential nitration, the non-volatile 2,4-dinitroanisole being quantitatively formed under conditions which left unchanged the *p*-chloronitrobenzene, which is volatile in steam.

Since this investigation Dey, Govindachari, and Venkatakrishna Udupa (1945) claim that a 96.6 per cent. yield of *p*-nitroanisole is obtained from *p*-chloronitrobenzene and 1.37N methanolic sodium hydroxide by stirring and heating for 15 hours with manganese dioxide as catalyst, using an amount equal to half the weight of the *p*-chloronitrobenzene. Lead dioxide has been shown to be equally effective as a catalyst (Dey 1948).

Reduction of p-Nitroanisole to p-Anisidine

Most of the common reducing agents have been used to reduce *p*-nitroanisole to *p*-anisidine (VI). Iron and hydrochloric acid is a very cheap reducing agent, but whilst the reduction proceeds very well in alcoholic solution, a considerable loss of *p*-anisidine occurs during isolation. Reduction with hydrogen using Raney nickel as catalyst was found to be a simpler method, and on a small scale a 90 per cent. yield of pure *p*-anisidine was obtained. Since the *p*-anisidine is produced in an atmosphere of hydrogen, no colouring matter is formed during the reduction, as occurs with other methods.

(c) 5-Diethylamino-2-aminopentane

An entirely new synthesis of 5-diethylamino-2-aminopentane (II) was begun in which the starting product was 1-chloro-3-bromopropane obtainable from propane-1,3-diol (which is a by-product in the production of glycerol). 1-Chloro-3-bromopropane readily gave, after Allen (1928), 1-chloro-3-cyanopropane which interacted with diethylamine to give a 75 per cent. yield of 1-diethylamino-3-cyanopropane. This latter should be convertible into II.

The investigation was discontinued when information was received that in the United States of America 2-diethylaminoethanol (prepared from diethylamine and ethylene oxide) was available technically as an intermediate for the synthesis of II.

III. EXPERIMENTAL

All melting points are uncorrected.

(a) *Preparation of 2,4-Dichlorobenzoic Acid (V)*

(i) *2-Chloro-*p*-toluenesulphonyl Chloride (VIII)*

(1) *Using Antimony Trichloride as Catalyst.* Pure *p*-toluenesulphonyl chloride (150 g.), recrystallized from light petroleum, b.p. 80–100 °C., and melting at 68.5 °C., and antimony trichloride (2 g)* were heated in an oil-bath at 70–80 °C. and treated with a stream of chlorine for 3 hours until the increase in weight was 28 g.

(The increase in weight of 28 g. is at first sight rather greater than the theoretical 27.7 g. (27.1 g. for the desired chlorination and 0.6 g. for the conversion of SbCl₃ into SbCl₅). Actually considerable under-chlorination has been effected, because several grams of chlorine and hydrogen chloride remain dissolved in the sulphonyl chloride at 70–80 °C. Since over-chlorination can readily take place (Davies 1921) and this would result in the eventual formation of trichlorobenzoic acids perhaps not easily separated from V, under-chlorination, as here, is desirable.)

While still warm, the product was freed from antimony† by washing three times with its own volume of concentrated hydrochloric acid. It was then stirred with warm water for a few minutes, extracted with benzene (500 ml.), the extract washed with dilute sodium hydroxide solution followed by water, and finally dried over calcium chloride. The benzene was removed by distillation and the residue distilled at reduced pressure giving the following fractions :

Fraction (i) 145–155 °C./19 mm. (33.5 g.)—mostly unchanged *p*-toluenesulphonyl chloride.

Fraction (ii) 155–156 °C./19 mm. (129.5 g.)—pure 2-chloro-*p*-toluenesulphonyl chloride (VIII).

Fraction (iii) 5 g. (approx.)—residue.

Yield of VIII, 72%. When fraction (i) is added to the next batch of VII, the yield of VIII is 94% (cf. McMaster and Carol 1931).

(2) *Using Thallous Chloride as Catalyst.*—Of the various catalysts it is found that iodine, though active, is, like antimony trichloride, difficult to remove completely by distillation. Thorium chloride is slightly reactive, whereas vanadium, cadmium, and manganous chlorides are inactive. Aluminium, ferric, and molybdenum chlorides are unsuitable as they produce, doubtless by a modified Friedel Craft reaction, hydrogen chloride and viscous material when warmed with *p*-toluenesulphonyl chloride (VII). Thallous chloride is insoluble in VII but when a mechanical stirrer is used, has about one-third of the activity of antimony trichloride.

p-Toluenesulphonyl chloride (100 g.) (recrystallized as above) was stirred with powdered thallous chloride (2.5 g.) (B.D.H.) in an oil-bath at 75–80 °C. The gas outlet was protected with a calcium chloride tube, and the mechanical stirrer operated through a mercury seal. Chlorine was introduced until the increase in weight was 20.5 g. (4½ hours) (see note to method 1). The reaction mixture was allowed to stand overnight, the clear yellow oil siphoned from the thallous chloride which had settled, and distilled at reduced pressure. The following fractions were obtained :

Fraction (i) 140–147 °C./14 mm. (2.6 g.)—unchanged *p*-toluenesulphonyl chloride (VII).

Fraction (ii) 147–150 °C./14 mm. (164 g.)—2-chloro-*p*-toluenesulphonyl chloride (VIII).

Fraction (iii) (8 g.)—residue (VIII).

Yield of VIII, 92%.

* A larger amount of antimony trichloride causes faster chlorination and this is not harmful.

† The complete removal of antimony is most important, otherwise the high temperature chlorination in the side-chain yields, in addition to 2,4-dichlorobenzotrichloride, pentachloro-toluenes and much non-volatile material.

(ii) *2,4-Dichlorobenzotrichloride (IX)*

2-Chloro-*p*-toluenesulphonyl chloride (102 g.) was heated in a long-necked flask in an oil-bath at 220–230 °C. A rapid stream of chlorine was passed through for 13½ hours. (This was about 1½ hours longer than the time required to remove from the exit gases the last traces of what seems to be sulphuryl chloride as shown by reaction with barium chloride to precipitate barium sulphate.) The resulting clear, amber coloured liquid was transferred, while still hot, to a distillation flask and distilled under reduced pressure. The following fractions were obtained :

Fraction (i) 115–143 °C./12 mm. (8 g.)—mostly 2,4-dichlorobenzal chloride.

Fraction (ii) 143–148 °C./12 mm. (102 g.)—2,4-dichlorobenzotrichloride (IX).

Fraction (iii) (7 g.)—residue of a dark red oil which crystallized on standing and was probably nearly all IX.

Fraction (ii) (IX) was a clear, colourless oil which solidified on standing, m.p. 37–50 °C. Yield 93%.

Pure 2,4-dichlorobenzotrichloride has b.p. 154 °C./19 mm., and m.p. 50 °C. (from light petroleum).

Found by alkaline hydrolysis : reactive Cl, 40.0%.

Calculated for $C_7H_3Cl_5$: Cl, 40.3%.

(Note.—Ultraviolet light in glass apparatus accelerates the chlorination of the side-chain to a slight extent.)

(iii) *2,4-Dichlorobenzoic Acid (V)*

(1) *Alkaline Hydrolysis*.—2,4-Dichlorobenzotrichloride (10 g. of fraction (ii)) was refluxed and stirred with a solution of sodium carbonate (12 g.) in water (150 ml.) for 6 hours. The solution was cooled and filtered, then acidified with hydrochloric acid. The precipitated 2,4-dichlorobenzoic acid was washed with water and dried in a steam oven overnight, m.p. 163 °C. Yield 7.20 g. (theory 7.208 g.).

(2) *Acid Hydrolysis*.—IX (40 g. of fraction (ii)) was stirred with fuming sulphuric acid (160 g. of $d_{20} = 1.870$; 4% SO_3) at 55–60 °C. for 2 hours. The dark red product was run into crushed ice (700 g.) and the white precipitate extracted with a hot solution of sodium carbonate (25 g.) in water. The extract was filtered from the undissolved 2,4-dichlorobenzaldehyde impurity, and the filtrate acidified with concentrated hydrochloric acid. The precipitate (V) was collected and dried as above. Yield 25.5 g. (89%). (Less than 1 g. of the aldehyde impurity was obtained.)

The overall yield for the three processes (i), (ii), and (iii) is over 80%.

(iv) *Derivatives of 2,4-dichlorobenzaldehyde*

2,4-Dichlorobenzotrichloride produced as described contained a small quantity of 2,4-dichlorobenzal chloride and on hydrolysis this was converted to 2,4-dichlorobenzaldehyde. This substance crystallized from ethyl acetate in white needles, m.p. 72 °C., and was characterized by the following derivatives :

p-Nitrophenylhydrazone.—Orange needles from ethyl acetate, m.p. 256–257 °C.

Found : N, 14.0%.

Calculated for $C_{13}H_9O_2N_3Cl_2$: N, 13.6%.

2,4-Dinitrophenylhydrazone.—A solution of 2,4-dinitrophenylhydrazine (1 g.) in concentrated sulphuric acid (2 ml.) was taken up in boiling alcohol (15 ml.). 5 ml. of this solution was added warm to an alcoholic solution of 2,4-dichlorobenzaldehyde (0.2 g.—slight excess). The precipitated 2,4-dinitrophenylhydrazone was recrystallized from toluene giving yellow crystals, m.p. 224–225 °C.

Found : N, 16.2%.

Calculated for $C_{13}H_8O_4N_4Cl_2$: N, 16.2%.

The semicarbazone, from alcohol, had m.p. 247–248 °C.

* This method of hydrolysis was that used by Erdmann and Schweten (1890) for the hydrolysis of 2,4-dichlorobenzal chloride to the corresponding aldehyde.

(b) *Preparation of p-Nitroanisole*

Sodium hydroxide (21 g., 96%) was added to methanol (200 ml.) and heated under reflux on the water-bath until most of it had dissolved. To the boiling solution was added *p*-chloronitrobenzene (79 g.), the mixture was then shaken, and refluxed for 12 hours. To the orange-red solution was then added sodium hydroxide (2.1 g., 96% = 10% of theory), a large amount (160 ml.) of the methanol removed (40 minutes), and the mixture refluxed for a further 3 hours 20 minutes (total time of heating was 16 hours). The hot mixture was poured into a large volume of cold water, heated to approximately 70 °C., and the *p*-nitroanisole then allowed to solidify, the mixture being stirred continually so that the *p*-nitroanisole solidified in small pellets. The yellow solid was filtered, washed several times with water to remove alkali and the soluble sodium *p*-nitrophenate formed during the reaction, allowed to drain, pressed, and finally dried under a high vacuum. The dry product was fractionally distilled under reduced pressure through a Vigreux column.

The following fractions were obtained:

Fraction (i) 144–151 °C./19 mm. (8.7 g.)—a pale yellow oil—unchanged *p*-chloronitrobenzene.

Fraction (ii) 155–156.5 °C./19.5 mm. (mostly at 155 °C.) (60.1 g.)—pure *p*-nitroanisole.

Fraction (iii) (3.3 g.)—dark red-brown residue—mostly *pp'*-dichloroazoxybenzene.

Yield of *p*-nitroanisole 78.5%, m.p. 51–52 °C.

The percentage yield for continuous work, where fraction (i) is added to the next batch of *p*-chloronitrobenzene, is about 85%.

(c) *Preparation of 1-Diethylamino-3-cyanopropane*

Powdered sodium iodide (0.1 g.) was added to 1-chloro-3-cyanopropane (30.5 g.) and heated (under an efficient reflux condenser) on the water-bath for half an hour. The flask was cooled, a soda lime tube was fitted to the top of the condenser, and anhydrous sodium carbonate (16 g.) added. Diethylamino (25 g., 16% excess of theory) was then added and the mixture refluxed on the water-bath for 19 hours. The contents of the flask were cooled, acidified with hydrochloric acid, and the oily layer of unchanged nitrile extracted with ether (100 ml.). The acidified solution was cooled in ice, and treated with 40% sodium hydroxide solution, when the 1-diethylamino-3-cyanopropane separated as an oil on the surface of the solution. The oil was extracted with six portions of ether, the extract dried over magnesium carbonate, and the ether removed by distillation. The residual oil was distilled under reduced pressure. Yield 31 g. (75%), b.p. 112–113 °C./40 mm. Refractive index $n_D^{22} = 1.4354$.

Found: N, 19.9%.

Calculated for $C_8H_{16}N_2$: N, 20.0%.

IV. ACKNOWLEDGMENTS

The authors are indebted to Mr. A. G. Dobson who carried out some preliminary investigations in the preparation of 2,4-dichlorobenzoic acid; to Mr. W. R. C. Stevenson for the microanalyses; to the Shell Company of Australia Ltd. for supplies of 1-chloro-3-bromopropane; and to J. Kitchen and Sons Pty. Ltd. for locally-produced 1,3-dihydroxypropane.

V. REFERENCES

- ALLEN, C. F. H. (1928).—“Organic Syntheses,” Vol. 8, p. 52. (John Wiley & Sons; New York); (Coll. Vol. 1, p. 156, 2nd Ed. (1944).)
- DAVIES, W. (1921).—*J. Chem. Soc.* **119**: 860.
- DAVIES, W., and DICK, J. H. (1932).—*J. Chem. Soc.* **1932**: 2045.
- DAVIES, W., and DOWN, R. A. R. (1929).—*J. Chem. Soc.* **1929**: 586.
- DEY, B. B. (1948).—Indian Pat. 33,473 (*Chem. Abstr.* **43**: 687 (1949)).
- DEY, B. B., GOVINDACHARI, T. R., and VENKATAKRISHNA UDUPA, H. (1945).—*J. Sci. Industr. Res. India* **4**: 374.

- D. R. PAT. 133,000 (1901).—*Friedlander* **6** : 53 (1904). (*Chem. Zbl.* **1902** II : 313 (1902).)
D. R. PAT. 145,908 (1902).—*Friedlander* **7** : 467 (1905). (*Chem. Zbl.* **1903** II : 1099 (1903).)
D. R. PAT. 234,290 (1909).—*Friedlander* **10** : 116 (1913). (*Chem. Zbl.* **1911** I : 1567 (1911).)
D. R. PAT. 639,578 (1935).—*Friedlander* **23** : 226 (1940). (*Chem. Zbl.* **1937** I : 2025 (1937).)
ERDMANN, H., and SCHWECTEN, E. (1890).—*Liebigs Ann.* **260** : 68.
HOLLEMAN, A. F., and DE MOOY, W. J. (1915).—*Rec. Trav. Chim. Pays-Bas* **35** : 17.
MCMASTER, L., and CAROL, J. (1931).—*Industr. Engng. Chem.* **23** : 218.
RARICK, M. J., BREWSTER, R. Q., and DAINS, F. B. (1933).—*J. Amer. Chem. Soc.* **55** : 1289.
RICHARDSON, D. H. (1926).—*J. Chem. Soc.* **1926** : 522.
U.S. PAT. 1,875,916 (1926).—*Chem. Zbl.* **1933** I : 675 (1933).
WISELOGLE, F. Y. (1946).—“A Survey of Antimalarial Drugs 1941-1945.” Vol. 1. p. 363 ;
Vol. 2. p. 1344. (J. W. Edwards : Ann Arbor, Michigan.)

THE DEMETHYLATION OF METHOXYACRIDONES

By G. K. HUGHES,* N. K. MATHESON,* A. T. NORMAN,* and E. RITCHIE*

[Manuscript received September 14, 1951]

Summary

4-Methoxy-10-methylacridone demethylates more readily than the 1-, 2-, or 3-isomers but not so readily as the acridone alkaloids. 4-Hydroxy-10-methylacridone resembles the nor-alkaloids but its isomers have distinctly different properties.

I. INTRODUCTION

An important and characteristic reaction of the acridone alkaloids is their ready conversion to nor-derivatives under mild acidic conditions (Brown *et al.* 1949; Crow and Price 1949*a*; Hughes and Neill 1949). Although it has been shown that in this reaction a methoxyl group in the 4-position is converted to a hydroxyl group (Crow and Price 1949*b*; Brown and Lahey 1950), it was clearly desirable to investigate the reaction further with simpler compounds, to determine the factors influencing it, and to compare the properties of hydroxy-10-methylacridones of known constitution with those of the nor-alkaloids.

Accordingly 1-, 2-, 3-, and 4-methoxy-10-methylacridones were prepared and their demethylation examined. The experiments were carried out under the following conditions:

- (a) Boiling with 1N aqueous ethanolic hydrochloric acid for 1 hour.
- (b) Boiling with 2N aqueous ethanolic hydrochloric acid for 8 hours.
- (c) Boiling with 4N aqueous ethanolic hydrochloric acid for 8 hours.
- (d) Heating the base hydrochloride at 200 °C. for 5 minutes.
- (e) Boiling with concentrated hydrobromic acid for 2 hours.

The conditions of (a) which were successful with melicopicine, melicopidine, melicopine (Crow and Price 1949*a*), and evoxanthine (Hughes and Neill 1949) affected none of the substances. Under conditions (b) and (c) 4-methoxy-10-methylacridone gave 4-hydroxy-10-methylacridone in yields of 25 and 10 per cent. (approx.) respectively, but the other substances were recovered unchanged. The 1- and 2-isomers were unaffected by condition (d), the 3-isomer gave a trace of a phenolic substance and was largely recovered unchanged, but the 4-isomer gave an excellent yield of the hydroxy-compound. All substances were completely demethylated by condition (e) to the corresponding hydroxy-10-methylacridones.

A marked difference was observed between the properties of 4-hydroxy-10-methylacridone and its isomers. The latter had no true melting points but gradually charred at high temperatures, were soluble in dilute sodium hydroxide

* School of Chemistry, University of Sydney.

to form orange solutions, had a green or blue fluorescence in alcoholic solution, and gave faint yellow ferric tests. The former, on the other hand, resembled the nor-alkaloids. It melted sharply at 190 °C., was insoluble in boiling dilute sodium hydroxide, showed no fluorescence in alcoholic solution, and gave a deep green ferric test.

Several other aspects of the demethylation reaction were also investigated. Melicopicine and 2,3,4-trimethoxy-10-methylacridone (Hughes, Neill, and Ritchie 1950) demethylate more readily than does 4-methoxy-10-methylacridone. It has now been found that *1,4-dimethoxy-10-methylacridone* occupies an intermediate position giving the *nor-derivatives* in 40 per cent. yield (approx.) under conditions (b). When the nitrogen atom was not methylated, demethylation was more difficult since 4-methoxy- and *1,4-dimethoxyacridones* were unaffected by conditions (a), (b), and (c). However, demethylation of both substances occurred on fusing their hydrochlorides, but demethylation of 1-, 2-, and 3-methoxyacridones was not effected under these conditions. In order to determine the behaviour of a 3,4-methylenedioxy-group, *1-methoxy-3,4-methylenedioxy-10-methylacridone* (rather than the inaccessible 3,4-methylenedioxy-10-methylacridone) was synthesized and subjected to conditions (a), (b), (c), and (d). It was either recovered unchanged or converted to dark amorphous materials.

II. THE MECHANISM OF DEMETHYLATION

Besides confirming the easy demethylation of the 4-methoxy-group in the 10-methylacridones the above results indicate that (i) the ease of demethylation is decreased in the corresponding acridones, (ii) the rate is increased where there are other methoxy- and/or methylenedioxy-groups present in the same nucleus, and (iii) the rate varies with the relative proportions of alcohol and acid.

Crow and Price (1949b) have tentatively suggested that the formation of the nor-alkaloids is associated with their high degree of stability due to hydrogen bonding. This argument can be extended to explain the mechanism of the reaction. It is generally agreed that the initial step in demethylation of ethers is the addition of a proton. In the case of the methoxyacridones this is most likely to occur at the carbonyl oxygen (cf. the formation of pyrylium salts). Now if a 4-methoxy-group is present in the molecule, the proton addition product will be stabilized by hydrogen bonding (see Fig. 1).

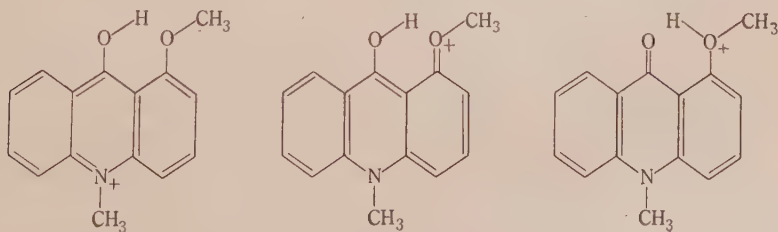


Fig. 1

This increase in stability will lower the activation energy necessary for the addition of a proton compared with that required when the methoxy-group is located elsewhere. Also if the carbonyl group is the most electronegative centre in the molecule, it further means that the addition of a second proton required for the demethylation of any of the 1-, 2-, or 3-methoxy-group will be hindered by a general inductive deactivation.

If now a methyl group is lost from the 4-methoxy-derivative, there is a further increase in stability as the molecule is now neutral and a stronger hydrogen bond is formed, a situation only possible where the methoxy-group is adjacent to the carbonyl group.

A similar situation will occur with the methoxyacridones. However, when the proton addition complex is formed, the further possibility arises that a proton can be lost from the nitrogen to give the tautomer of the original acridone (Fig. 2). This would lessen the strain on the ethereal bond and thus

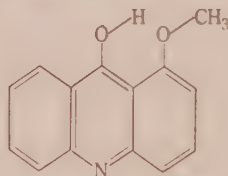


Fig. 2

make demethylation more difficult than in the 10-methylacridones. Nevertheless it would be expected that the ease of formation of the adduct in the above case would make demethylation easier than with the other monomethoxy-derivatives.

The above cases are similar to that of the demethylation of visnagin by boiling dilute hydrochloric acid (Schönberg and Badran 1951). It is conceivable that mild conditions rather than the conventional hydrobromic acid or hydriodic acid would be more suitable for many demethylations where the formation of a hydrogen bond is possible.

The incidental points concerning change of rate of demethylation in the polymethoxy-derivatives and the variation with concentration of alcohol will not be discussed until more information has been accumulated.

III. THE PREPARATION OF 10-METHYLACRIDONES

The cyclization of diphenylamine-2-carboxylic acids by phosphorus oxychloride to 5-chloroacridines followed by acid hydrolysis forms an easy route to acridones, but 10-methylacridones are not so readily prepared. In seeking a general method for the synthesis of methoxylated derivatives of 10-methylacridone, attention was first directed to the preparation of the monomethoxy-derivatives, and the following reactions were considered:

(1) Alkylation of the potassium salt of the acridone prepared by evaporating to dryness the solution of the acridone and potassium hydroxide in methanol (Graebe and Lagodzinski 1892; Drummond and Lahey 1949).

(2) Treatment of the methiodide of the 5-aminoacridine with aqueous sodium carbonate (Albert and Ritchie 1943).

(3) Alkaline hydrolysis of the metho-*p*-toluenesulphonate of the 5-phenoxy-acridine (Drozdov 1938).

(4) Treatment of the methosulphate of the 5-chloroacridine with aqueous sodium hydroxide (Glen and Nitzsche 1939).

(5) Rearrangement of the 5-methoxyacridine by heating with methyl iodide at 100 °C. (Hughes, Neill, and Ritchie 1950).

(6) Oxidation of the methosulphate of the acridine with alkaline ferri cyanide (Decker 1892; Hughes and Ritchie 1951).

Glen and Nitzsche (1939) prepared 4-methoxy-10-methylacridine in unstated yield by method (1), but were evidently unable to obtain a pure sample of the 1-isomer, which they described as an oil. Repetition of this preparation yielded either oils or unchanged acridone. In addition the slight solubility of many acridones in methanol rendered the method unattractive for the preparation of reasonable quantities of material. Methods (2), (3), and (4) are closely related and may reasonably be expected to have similar limitations. Although a good yield of 10-methylacridone was obtained by Albert and Ritchie (1943) by method (2), it failed with the tri- and tetramethoxy-derivatives, yielding only the acridones on hydrolysis with sodium carbonate (Hughes, Neill, and Ritchie 1950), and similar results were obtained in attempts to prepare 1- and 4-methoxy-10-methylacridones. Method (3) also gave only the acridones in these two cases although it succeeded with the 3-methoxy-derivative (Drozdov 1938). Glen and Nitzsche (1939) obtained 2- and 3-methoxy-10-methylacridones in unstated yield by method (4), but found it unsuitable for the preparation of the 1- and 4-isomers.

Methods (5) and (6), however, were found to be quite general and each of the methoxy-10-methylacridones was prepared in satisfactory yields. In addition 1,4-dimethoxy- and 1-methoxy-3,4-methylenedioxy-10-methylacridones were prepared by method (5) from the respective 5-chloroacridines.

2,5-Dimethoxydiphenylamine-2'-carboxylic acid was prepared in the usual manner from bromoquinol dimethyl ether and anthranilic acid. 2-Methoxy-4,5-methylenedioxydiphenylamine-2'-carboxylic acid was synthesized from anthranilic acid and 5-bromo-4-methoxy-1,2-methylenedioxybenzene obtained through the intermediates, piperonal, 4-hydroxy-1,2-methylenedioxybenzene, and 4-methoxy-1,2-methylenedioxybenzene.

For method (6) the methoxyacridines were required. An easy and general route to these appeared to be the dehydrogenation of 1,2,3,4-tetrahydroacridines prepared from the arylamine and 2-formylcyclohexanone by the method of Petrow (1942). Tetrahydroacridine and its 7- and 8-methoxy-derivatives were smoothly dehydrogenated by palladium-charcoal in boiling diphenyl ether to acridine and 3- and 2-methoxyacridines respectively, but the 9-methoxy-derivative gave 1-methoxyacridine in very poor yield. In the reaction between *m*-anisidine and 2-formylcyclohexanone only 8-methoxy-1,2,3,4-tetrahydroacridine could be isolated, no trace of the 6-isomer being detected. A similar result was obtained by Petrow (1942) with *m*-toluidine. The required 1- and 4-methoxyacridines were eventually obtained in good yield from the corresponding

5-chloroacridines by condensing them with *p*-toluenesulphonhydrazide and decomposing the adducts with alkali according to Albert and Royer (1949). Oxidation of the *methosulphates* of the methoxyacridines to the methoxy-10-methylacridones by hot alkaline ferricyanide proceeded in nearly quantitative yields.

IV. EXPERIMENTAL

(Light petroleum refers to the fraction of b.p. 60–80 °C.)

(a) 3-Methoxydiphenylamine-2'-carboxylic Acid

The preparation of this acid has been very briefly described by Ullmann (1907) and a confused account has been given by Lehmstedt and Schrader (1937). Both record its m.p. as 132 °C. The substance was prepared as follows. To a hot solution of *o*-chlorobenzoic acid (30 g.) in amyl alcohol (200 cc.) was gradually added dry powdered potassium carbonate (30 g.), then when the reaction had subsided, *m*-anisidine (25 g.) and copper powder (1 g.). After refluxing for 1 hour, the mixture was steam distilled, the residual liquid treated with charcoal, filtered, and the hot filtrate (about 1000 cc.) acidified. The granular precipitate was collected, washed with boiling water (about 1000 cc.), dried, and recrystallized from a small volume of benzene with the aid of charcoal, giving nearly colourless material (25 g.; 50% yield), m.p. 134–135 °C. The pure substance separated from benzene in colourless flattened needles, m.p. 137 °C.

Found: C, 69.3; H, 5.5%.

Calculated for $C_{14}H_{13}O_3N$: C, 69.1; H, 5.4%.

(b) 2- and 4-Methoxy-5-chloroacridines

Cyclization of the above acid (75 g.) by phosphorus oxychloride and separation of the products by the method of Albert (1951, p. 40) gave 2-methoxy-5-chloroacridine (42 g.) and eventually 4-methoxyacridone (12 g.). Two treatments with acid and a recrystallization from nitrobenzene were required before a satisfactory m.p. (347 °C. decomp. with previous sintering) for the latter substance was reached, and the amount obtained was nearly twice that to be expected according to Albert. Phosphorus oxychloride converted it quantitatively to 4-methoxy-5-chloroacridine.

(c) 1- and 3-Methoxy-5-chloroacridines

These substances were prepared according to Gleu and Nitzsche (1939).

(d) 1-, 2-, 3-, and 4-Methoxyacridones

The corresponding chloro-compounds were hydrolysed by acid (Gleu and Nitzsche 1939).

(e) 2,5-Dimethoxydiphenylamine-2'-carboxylic Acid

After several preliminary experiments in which the time, solvent, temperature, and catalyst were varied, the following method was evolved. A mixture of anthranilic acid (12.7 g.), 2-bromo-1,4-dimethoxybenzene (20 g.), dry powdered potassium carbonate (13 g.), cuprous chloride (6 g.), and cyclohexanone (70 cc.) was refluxed for 4 hours and the product isolated as above. By recrystallization from benzene, pale yellow needles (11.5 g.; 45% yield), m.p. 163 °C. were eventually obtained.

Found: N, 5.2%.

Calculated for $C_{16}H_{15}O_4N$: N, 5.1%.

(f) 1,4-Dimethoxy-5-chloroacridine

Cyclization of the above acid by phosphorus oxychloride and isolation of the product by the chloroform ammonia method in the usual way gave a nearly quantitative yield. The substance crystallized from benzene in greenish yellow needles, m.p. 142 °C.

Found: N, 5.3%.

Calculated for $C_{16}H_{13}O_2NCl$: N, 5.3%.

(g) 1,4-Dimethoxyacridone

The chloro-compound (1.2 g.) was hydrolysed by heating it with 5N hydrochloric acid (50 cc.) for 2 hours on the water-bath. After basifying with ammonia, the product was collected and recrystallized several times from alcohol. It formed small yellow needles (0.6 g.; 55% yield), m.p. 221 °C.

Found : CH_3O , 24.0%.

Calculated for $\text{C}_{15}\text{H}_{13}\text{O}_3\text{N}$: $2 \times \text{CH}_3\text{O}$, 24.3%.

(h) 1-, 2-, 3-, and 4-Methoxy- and 1,4-Dimethoxy-10-methylacridones

Conversion of the 5-chloroacridines by methanolic sodium methoxide to the 5-methoxyacridines and rearrangement of the latter to the 10-methylacridones by heating with methyl iodide was effected as previously described (Hughes, Neill, and Ritchie 1950). Information concerning the products is summarized in Table 1.

TABLE 1
ALKOXY-10-METHYLACRIDONES

Substituent (and Formula)	Melting Point (°C.)		Habit and Colour	Solvent	Yield (%).	Analysis			
	This Work	Gleu and Nitzsche (1939)				Found		Calculated	
						C (%)	H (%)	C (%)	H (%)
1-Methoxy (C ₁₅ H ₁₃ O ₂ N)	91	Oil	Pale yellow blades	cycloHexane	45	75.2	5.7	75.3	5.5
2-Methoxy (C ₁₅ H ₁₃ O ₂ N)	185	185	Very pale yellow needles	Alcohol	55	75.1	5.5	75.3	5.5
3-Methoxy (C ₁₅ H ₁₃ O ₂ N)	143*	147-148	Yellow needles	Aqueous al- cohol or benzene-light petroleum	55	75.1	5.6	75.3	5.5
4-Methoxy (C ₁₅ H ₁₃ O ₂ N)	165	164	Cream needles	Benzene	50	75.4	5.4	75.3	5.5
1,4-Dimethoxy (C ₁₆ H ₁₅ O ₃ N)	80	—	Yellow needles	Aqueous alcohol	40	71.6	5.6	71.4	5.6

* Not altered by further recrystallization.

(i) 4-Methoxy-1,2-methylenedioxybenzene

Dimethyl sulphate (17.5 cc.) was gradually added with shaking to a solution of 4-hydroxy-1,2-methylenedioxybenzene (17 g.), prepared from piperonal (Böeseke, Cohen, and Kip 1936), in aqueous potassium hydroxide (50 cc. of 30%). After heating the mixture on the water-bath for 1 hour, it was cooled, extracted with ether, and the product isolated by distillation. It formed a colourless oil, b.p. 134-136 °C./40 mm. (15 g.; 80% yield).

Found : C, 62.8; H, 5.1%.

Calculated for $\text{C}_8\text{H}_8\text{O}_3$: C, 63.1; H, 5.3%.

(j) *5-Bromo-4-methoxy-1,2-methylenedioxybenzene*

(i) A solution of bromine (14 g.) in chloroform (50 cc.) was added gradually with shaking to a solution of the ether (13.3 g.) in chloroform (50 cc.). The colour of the bromine was rapidly discharged and hydrogen bromide was evolved. The reaction mixture was washed with water, dried, evaporated, and the residue distilled. The product (18 g.; 90% yield) collected at 112–114 °C./1 mm. quickly solidified. Recrystallization from aqueous alcohol gave colourless needles, m.p. 82 °C.

Found: C, 41.5; H, 3.1; Br, 34.3%.

Calculated for $C_8H_7O_3Br$: C, 41.6; H, 3.1; Br, 34.6%.

(ii) Although there was little doubt that this substance had the structure assigned to it, proof was obtained by synthesizing it from 4-bromo-5-nitro-1,2-methylenedioxybenzene (Jones and Robinson 1917).

4-Bromo-5-amino-1,2-methylenedioxybenzene.—An intimate mixture of the nitro-compound (5 g.) with iron powder (3.2 g.) was added during 20 minutes to alcohol (30 cc.) and concentrated hydrochloric acid (1 cc.), stirred, and heated under reflux. After a further 2 hours, the mixture was filtered, the filtrate concentrated to small bulk and made strongly alkaline. The product (1.5 g.; 35% yield), m.p. 80 °C., isolated by steam distillation formed greyish plates.

Found: N, 6.6%.

Calculated for $C_7H_6O_2NBr$: N, 6.5%.

Attempts to effect reduction by tin or stannous chloride and hydrochloric acid gave dark tars.

The *acetyl derivative* crystallized from aqueous alcohol in colourless needles, m.p. 148 °C.

Found: N, 5.3%.

Calculated for $C_9H_8O_3NBr$: N, 5.4%.

5-Bromo-4-methoxy-1,2-methylenedioxybenzene.—To a solution of the above amine (5 g.) in glacial acetic acid (75 cc.) was added sulphuric acid (20 cc. of 10N) and the mixture cooled to 0 °C. Sodium nitrite (1.7 g.) in water (10 cc.) was slowly added during 30 minutes with stirring and then the resulting diazo-solution poured into water (3000 cc.). Decomposition of the diazonium salt occurred slowly when this solution was stirred in sunlight. When evolution of nitrogen had ceased, the mixture was filtered from a dark tar, the filtrate nearly neutralized with sodium carbonate, and then exhausted with ether. The extract was concentrated, extracted with aqueous potassium hydroxide (50 cc. of 30%), and the alkaline liquid shaken immediately with dimethyl sulphate (8 cc.). After warming on the water-bath for an hour, the product (1.1 g.; 20% yield) was recovered by steam distillation. Recrystallization from aqueous alcohol gave colourless needles, m.p. and mixed m.p. 82 °C.

Other methods of diazotization and decomposition of the diazonium salt gave only minute traces of the product.

(k) *2-Methoxy-4,5-methylenedioxydiphenylamine-2'-carboxylic Acid*

A mixture of anthranilic acid (30 g.), the above bromo-compound (35 g.), dry powdered potassium carbonate (30 g.), copper powder (1 g.), and nitrobenzene (200 cc.) was heated at 180–185 °C. for 2 hours with frequent shaking. The crude product isolated in the usual way was purified by recrystallization from alcohol. It formed pale yellow prisms (15 g.; 35% yield), m.p. 220 °C.

Found: C, 62.6; H, 4.5; N, 4.8%.

Calculated for $C_{16}H_{13}O_5N$: C, 62.7; H, 4.6; N, 4.9%.

(l) *1-Methoxy-3,4-methylenedioxy-5-chloroacridine*

This substance obtained in nearly quantitative yield in the usual manner, crystallized from xylene in bright red needles, m.p. 240 °C. (decomp.). Its solution in phosphorus oxychloride was dark blue.

Found: C, 62.8; H, 3.7; Cl, 12.3%.

Calculated for $C_{16}H_{10}O_3NCl$: C, 62.6; H, 3.5; Cl, 12.3%.

(m) 1,5-Dimethoxy-3,4-methylenedioxyacridine

The reaction between the chloroacridine and methanolic sodium methoxide was slow and incomplete under the usual conditions, but the following procedure gave good results.

Sodium (0.2 g.) was dissolved in methanol (15 cc.) and the solution added to a solution of the chloroacridine (1.5 g.) in hot xylene (40 cc.). After distilling off the methanol, the mixture was refluxed for 30 minutes and then filtered. The filtrate was cooled somewhat, and an equal volume of light petroleum added. On standing, the product (0.9 g.; 60% yield) slowly separated. Recrystallization from benzene gave orange needles, m.p. 187 °C.

Found: C, 67.7; H, 4.8%.

Calculated for $C_{18}H_{18}O_4N$: C, 67.8; H, 4.6%.

(n) 1-Methoxy-3,4-methylenedioxy-10-methylacridone

Rearrangement of the above substance (0.8 g.), by heating with methyl iodide (4 cc.) at 100 °C. for 4 hours in the usual manner, followed by recrystallization from benzene, gave glistening orange needles (0.6 g.; 75% yield), m.p. 164 °C.

Found: C, 67.7; H, 4.8%.

Calculated for $C_{18}H_{13}O_4N$: C, 67.8; H, 4.6%.

(o) 1-(m-Methoxyphenyliminomethyl)cyclohexan-2-one

m-Anisidine (17.5 g.) was added to a warm solution of formylcyclohexanone (18 g.) in alcohol (50 cc.). In a short time the product (25.1 g.; 75% yield) began to separate in small yellow irregular plates, m.p. 113 °C.

Found: N, 6.2%.

Calculated for $C_{14}H_{17}O_2N$: N, 6.1%.

(p) 8-Methoxy-1,2,3,4-tetrahydroacridine

A solution of the anil (23.6 g.; 1 mol.), *m*-anisidine hydrochloride (31.8 g.; 2 mol.), and anhydrous zinc chloride (13.6 g.; 1 mol.) in absolute alcohol (300 cc.) was refluxed for 12 hours. After cooling, the yellow crystalline precipitate, which had begun to separate after 1 hour, was collected and the filtrate evaporated to small bulk. The mixture was made strongly alkaline and extracted with ether. The extract was dried over sodium hydroxide, the ether removed, and the residue distilled under reduced pressure. A pale yellow oil (16.5 g.; 75% yield), b.p. 227–230 °C./20 mm., was obtained which solidified completely on cooling.

To determine whether the product was a mixture it was converted to its picrate which was then fractionally crystallized from acetone, but only one substance could be isolated. The substance regenerated from its picrate, crystallized from light petroleum in large colourless transparent tablets, m.p. 61 °C.

Found: N, 6.7%.

Calculated for $C_{14}H_{16}ON$: N, 6.6%.

It was very soluble in the usual organic solvents including light petroleum but it formed a sparingly soluble hydrochloride. Impure specimens showed a violet fluorescence in most solvents, which appeared to be due to traces of acid since the pure substance, which showed no fluorescence in organic solvents, also had a violet fluorescence in dilute aqueous acid.

The *picrate* crystallized from acetone in yellow elongated hexagons, m.p. 200–201 °C. (decomp.).

Found: N, 12.8%.

Calculated for $C_{20}H_{18}O_8N_4$: N, 12.7%.

(q) Dehydrogenation of 1,2,3,4-Tetrahydroacridines

Since dehydrogenation by heating with lead oxide, or zinc dust (Borsche 1910), or selenium (Petrow 1942) gave relatively low yields, other methods were investigated using tetrahydroacridine as a model substance. Heating with chloranil in xylene (Barclay and Campbell 1945)

led to rapid formation of intractable tars. Palladium-charcoal in boiling *p*-cymene or cyclohexanone was unsuccessful but by using boiling diphenyl ether or 1,2,4-trichlorobenzene good yields were obtained. The following procedure was adopted.

A solution of tetrahydroacridine (5 g.) in diphenyl ether (15 cc.) was refluxed with palladium-charcoal (1 g. of 10%) for 2 hours. After cooling, the mixture was diluted with ether, the catalyst filtered off, and the ether removed from the filtrate. The residue was extracted with boiling hydrochloric acid (2×100 cc. of N,5), the extract treated with charcoal, filtered, and basified. The mixture was extracted with benzene, the extract concentrated, and then passed through a short column of alumina. Evaporation of the benzene eluates gave nearly pure acridine (3.5 g.: 70% yield).

3-Methoxyacridine.—This substance was obtained from 7-methoxytetrahydroacridine (Petrow 1942) in 80% yield by the same procedure, except that more hydrochloric acid (500 cc.) was used to dissolve its sparingly soluble hydrochloride. The pure substance crystallized from light petroleum as orange-yellow needles, m.p. 105°C . Lehnstedt, Bruns, and Klee (1936) give m.p. 104°C .

Found: C, 80.1; H, 5.4%.

Calculated for $\text{C}_{14}\text{H}_{11}\text{ON}$: C, 80.4; H, 5.3%.

2-Methoxyacridine.—The crude product obtained in 65% yield was recrystallized from light petroleum. It formed very pale yellow needles, m.p. 91°C . Sherlin *et al.* (1938) quoted by Albert (1951, p. 137) give m.p. $89\text{--}90^{\circ}\text{C}$.

Found: C, 80.2; H, 5.4%.

1-Methoxyacridine.—The yield from 9-methoxytetrahydroacridine (Petrow 1942) was very low (15%) and purification was troublesome. The substance crystallized from cyclohexane in pale yellow opaque needles, m.p. 134°C . Albert (1951, p. 137) gives m.p. 134°C .

Found: C, 80.4; H, 5.4%.

(r) 4-Methoxyacridine

When 4-methoxy-5-chloroacridine (10 g.; 1 mol.) in chloroform (50 cc.) was added to *p*-toluenesulphonhydrazide (9 g.; 1.1 mol.) in chloroform (150 cc.) a mildly exothermic reaction occurred, the solution became very deep red and in about 15 minutes the adduct began to separate. The next day a rapid stream of dry hydrogen chloride was passed in for about 10 seconds and the mixture then allowed to stand for 2 days. The yellow-orange adduct (18.0 g.) was collected, washed thoroughly with chloroform, and air-dried.

To a solution of sodium hydroxide (40 g.) in water (300 cc.) and ethylene glycol (700 cc.), vigorously stirred and heated in a water-bath, the powdered adduct was added in one portion. After 3 hours, the mixture was cooled, diluted, and extracted with benzene. The benzene was evaporated, the residue extracted with hot hydrochloric acid (200 cc. = 50 cc. of N 2) and the extract after treatment with charcoal, filtered, basified, and shaken with benzene. The benzene solution was concentrated and then passed through a short column of alumina. From the benzene eluates a nearly pure product (7 g.; 80% overall yield) was obtained. The pure substance crystallized from light petroleum in small yellow clumps of plates, m.p. 122°C , not altered by further recrystallization. Sherlin *et al.* (1939), quoted by Albert (1951, p. 137), give m.p. $155\text{--}156^{\circ}\text{C}$, which is surely incorrect.

Found: C, 80.3; H, 5.3%.

Calculated for $\text{C}_{14}\text{H}_{11}\text{ON}$: C, 80.4; H, 5.3%.

(s) 1-Methoxyacridine

The formation of the adduct was slower, being allowed to proceed for 4 days, but otherwise the procedure was the same as above. The overall yield of crude product was 70%. The pure substance had m.p. and mixed m.p. 134°C .

(t) *Methosulphates of the Monomethoxyacridines*

A solution of the base (5 g.) and dimethyl sulphate (10 cc.) in pure dry benzene (50 cc.) was gently refluxed. Usually the product began to separate in a few minutes but refluxing was continued for 1 hour. After cooling, the product was collected and washed with benzene and acetone. The latter solvent removes traces of highly coloured impurities which are usually present unless quite pure starting materials are used. The yields were almost quantitative. The properties of the products are summarized in Table 2.

TABLE 2
METHOSULPHATES OF MONOMETHOXYACRIDINES

Sub- stituent	Melting Point (°C.)	Habit and Colour	Solvent	Found*		Remarks
				C (%)	H (%)	
1	170	Bright yellow needles	Absolute alcohol ethyl acetate	57.6	5.2	Hygroscopic; ther- mochromic becom- ing red below 100 °C.
2	260 Decomp. with previous charring	Fine yellow needles	Alcohol	57.2	5.2	Less soluble than its isomers
3	227 Decomp. with previous sintering	Very small yellow needles	Alcohol	57.2	5.0	
4	216–217 Decomp. with previous charring	Small orange needles	Absolute alcohol	57.5	5.1	Thermochromic be- coming red below 100 °C.

* Calculated for $C_{16}H_{17}O_5NS$: C, 57.3; H, 5.1%.

(u) *Oxidation of the Quaternary Salts*

Oxidation with hot alkaline ferricyanide and isolation of the product was effected as previously described (Hughes and Ritchie 1951). The volume of the reaction mixture was nearly trebled in the case of the 2-isomer because of the low solubility of the quaternary ferricyanide. The yield of crude but nearly pure product was almost quantitative in each case. The substances were respectively identical with those obtained by the alternative procedure.

Contrary to the report of Lehmstedt and Hundertmark (1931), 10-methylacridone may also be prepared in quantitative yield by this reaction. That the ferricyanide is involved in the oxidation is shown by its conversion to ferrocyanide.

(v) *Demethylation Experiments*

(i) The substance (0.5 g.) was refluxed for 1 hour with alcohol (27 cc.) and concentrated hydrochloric acid (3 cc.). The solution was then quickly evaporated to small bulk, made ammoniacal, and the product collected.

(ii) The substance was refluxed for 8 hours with alcohol (24 cc.) and concentrated hydrochloric acid (6 cc.), and the product isolated as above.

(iii) The substance was refluxed for 8 hours with alcohol (18 cc.) and concentrated hydrochloric acid (12 cc.).

TABLE 3
HYDROXYACRIDONES

Substituent (and Formula)	Melting Point (°C.)	Habit and Colour	Analysis			Remarks
			Found (%)	Calculated (%)	Calculated (%)	
			C	H	C H	
			(%)	(%)	(%) (%)	
1-Hydroxy-10-methyl ($C_{14}H_{11}O_2N$)	260 Decomp. with previous charring	Small brownish yellow prisms	74.7	5.0	74.6 4.9	Alcoholic solution fluoresces green; orange solution in aqueous sodium hydroxide; yellow ferric test
2-Hydroxy-10-methyl ($C_{14}H_{11}O_2N$)	300 Decomp. with previous charring	Pale yellow blades	74.3	5.0	74.6 4.9	Alcoholic solution fluoresces violet; orange solution in aqueous sodium hydroxide; yellow ferric test
3-Hydroxy-10-methyl ($C_{14}H_{11}O_2N$)	295 Decomp. with previous charring	Dull yellow blades	74.5	4.8	74.6 4.9	Alcoholic solution fluoresces blue-green; orange solution in aqueous sodium hydroxide; yellow ferric test
4-Hydroxy-10-methyl ($C_{14}H_{11}O_2N$)	190	Orange needles	74.5	4.9	74.6 4.9	Alcoholic solution non-fluorescent; insoluble in aqueous alkali; green ferric test
4-Hydroxy ($C_{13}H_9O_2N$)	320 Decomp.	Dull orange needles	73.7	4.2	73.9 4.3	Alcoholic solution non-fluorescent; insoluble in aqueous alkali; green ferric test
4-Hydroxy-1-methoxy ($C_{14}H_{11}O_3N$)	229	Orange plates	69.4	4.6	69.7 4.6	Alcoholic solution non-fluorescent; insoluble in aqueous alkali; green ferric test
4-Hydroxy-1-methoxy-10-methyl ($C_{15}H_{13}O_3N$)	136	Orange needles	70.5	5.0	70.6 5.1	Alcoholic solution non-fluorescent; insoluble in aqueous alkali; green ferric test

(iv) The hydrochloride of the base, prepared by adding a little concentrated hydrochloric acid to a solution of the base in acetone, was heated at 200 °C. until molten and then maintained at this temperature for 5 minutes. In the case of the acridones, the hydrochlorides were kept molten for 3 minutes by carefully heating with a free flame.

(v) The substance (0.5 g.) was refluxed for 2 hours with concentrated hydrobromic acid (10 cc.). After diluting, the mixture was made ammoniacal, warmed to coagulate the precipitate, then cooled and filtered.

The results of these experiments have been described above. The properties of the hydroxy-derivatives are summarized in Table 3. All were recrystallized from alcohol. The nor-derivatives were less soluble than the parent compounds, thus permitting a fairly ready separation.

V. ACKNOWLEDGMENTS

The authors are indebted to Mrs. E. Bielski for the analyses, to the Commonwealth Research Grants Committee, University of Sydney, for the award of a scholarship to one of them (A.T.N.), and to Mr. W. J. Herzberg for the preparation of 5-bromo-4-methoxy-1,2-methylenedioxybenzene from 4-bromo-5-nitro-1,2-methylenedioxybenzene.

VI. REFERENCES

- ALBERT, A. (1951).—"The Acridines." (Edward Arnold & Co.: London.)
ALBERT, A., and RITCHIE, B. (1943).—*J. Chem. Soc.* **1943**: 458.
ALBERT, A., and ROYER, R. (1949).—*J. Chem. Soc.* **1949**: 1148.
BARCLAY, B. M., and CAMPBELL, N. (1945).—*J. Chem. Soc.* **1945**: 530.
BÖESEKEN, J., COHEN, W. D., and KIP, C. J. (1936).—*Rec. Trav. Chim. Pays-Bas.* **55**: 815.
BORSCHKE, W. (1910).—*Liebigs Ann.* **377**: 70.
BROWN, R. D., DRUMMOND, L. J., LAHEY, F. N., and THOMAS, W. C. (1949).—*Aust. J. Sci. Res.* **A 2**: 622.
BROWN, R. D., and LAHEY, F. N. (1950).—*Aust. J. Sci. Res.* **A 3**: 593.
CROW, W. D., and PRICE, J. R. (1949a).—*Aust. J. Sci. Res.* **A 2**: 255.
CROW, W. D., and PRICE, J. R. (1949b).—*Aust. J. Sci. Res.* **A 2**: 282.
DECKER, H. (1892).—*J. prakt. Chem.* **45**: 161.
DROZDOV, N. S. (1938).—*Chem. Abstr.* **32**: 568.
DRUMMOND, L. J., and LAHEY, F. N. (1949).—*Aust. J. Sci. Res.* **A 2**: 630.
GLEU, K., and NITZSCHE, S. (1939).—*J. prakt. Chem.* **153**: 200.
GRAEBE, C., and LAGODZINSKI, K. (1892).—*Ber. dtsh. chem. Ges.* **25**: 1733.
HUGHES, G. K., and NEILL, K. G. (1949).—*Aust. J. Sci. Res.* **A 2**: 429.
HUGHES, G. K., NEILL, K. G., and RITCHIE, E. (1950).—*Aust. J. Sci. Res.* **A 3**: 497.
HUGHES, G. K., and RITCHIE, E. (1951).—*Aust. J. Sci. Res.* **A 4**: 423.
JONES, T. G. H., and ROBINSON, R. (1917).—*J. Chem. Soc.* **111**: 903.
LEHMSTEDT, K., BRUNS, W., and KLEE, H. (1936).—*Ber. dtsh. chem. Ges.* **69**: 2399.
LEHMSTEDT, K., and HUNDERTMARK, H. (1931).—*Ber. dtsh. chem. Ges.* **64**: 2386.
LEHMSTEDT, K., and SCHRADER, K. (1937).—*Ber dtsh. chem. Ges.* **70**: 838.
PETROW, V. A. (1942).—*J. Chem. Soc.* **1942**: 695.
ULLMANN, F. (1907).—*Liebigs Ann.* **355**: 312.
SCHÖNBERG, A., and BADRAN, N. (1951).—*J. Amer. Chem. Soc.* **73**: 2960.
SHERLIN, S. M., BRAZ, G. I., JAKUBOVICH, A. Y., VOROBEJEVA, E. I., and RABINOVICH, F. E. (1939).—*Chem. Abstr.* **33**: 1330.

SHORT COMMUNICATIONS

THE CALCULATION OF THE PROBABLE ERROR OF DETERMINATIONS OF LUNAR DAILY HARMONIC COMPONENT VARIATIONS IN GEOPHYSICAL DATA : A CORRECTION*

By SYDNEY CHAPMAN†

1. In a recent paper Tschu (1949) has set out in some detail the practical execution of the Chapman-Miller (1940) method for the determination of lunar and lunisolar daily variations in geophysical data.

Tschu (1949) indicated how the probable error of any such determination could be calculated, along lines developed by Miller (1934). This secondary but important part of Tschu's paper needs a correction, namely, the reversal of sign of the entries in the columns headed B_r and C_r of his Table 4.

TABLE 1

$\Delta a_2'$	$\Delta b_2'$	$\Delta a_2''$	$\Delta b_2''$	$\sqrt{\{(\Delta a_2'')^2 + (\Delta b_2'')^2\}}$
2.76	0.01	0.52	0.08	0.52
2.72	1.81	0.48	1.90	1.96
1.54	-0.33	-0.70	-0.24	0.75
1.96	-1.56	-0.28	-1.47	1.50
3.30	0.03	1.06	0.12	1.06
1.82	0.28	0.41	0.19	0.45
1.81	0.82	-0.43	0.91	0.96
2.25	-0.80	0.01	-0.71	0.71
2.55	0.58	0.31	0.67	0.71
2.42	-0.10	0.18	-0.01	0.18
1.81	0.13	-0.43	0.22	0.48
1.98	-1.41	-0.26	-1.32	1.34
26.93	-1.10			10.62
2.24	-0.09			$R' = 0.885$

$$L^* = \sqrt{(L^2 + V^2)} = 2.243$$

$$L^* R^* = 0.931$$

$$L^* R = 2.53$$

From Table 3, $P.E. R^*$ or Table 5, $P.E. R'$ (delete the one not used) = 0.0508, whence $P.E. = 0.045$ units = 6.0 microbars. $L_2 = 52.4 \pm 6.0$ microbars.

The error affects the part of his worked-out example that relates to the calculation of the probable error. The entries in Table 1 should replace those printed in the right-hand and lower portions of Tschu's Figure 6, p. 19.

* Manuscript received August 30, 1951.

† The Queen's College, Oxford.

It is satisfactory to find that the true probable error (6.0 microbars) of this determination (for the lunar atmospheric tide at Huancayo, Peru, in the 4 months May to August) is so much less than 16.3 microbars, as incorrectly calculated in Tschu's (1949) paper.

The error here corrected arose in connection with a change in the basis of selection (in the application of the Chapman-Miller method) of the groups of solar days at 24 epochs in the lunar month. Up to about 1938 the ordinal numbers 0, 1, 2, . . . of the groups represented the exact mean Greenwich solar hours nearest to the (Greenwich *transit-times*) of the (real) moon. These numbers *increased* steadily from one new moon (group 0) to the next. After about 1938 a revised plan (that described by Tschu 1949) was adopted, in which the group-numbers 0, 1, . . ., represented the value of μ for the days in each group, where μ denotes a number (see §4 of this communication) introduced by Schmidt (1928) and tabulated by Bartels and Fanslau (1937, 1938), which *decreases* from 24 to 0 from one new (mean) moon to the next; thus the order (in the lunar month) of the groups of days numbered 0, 1, 2, . . . is the reverse of that adopted in the earlier practice. The real importance of the change rested in the reference to the *mean* instead of the *real* moon. The reversal of order of the groups naturally necessitated changes in the details of the computations; the published version of the Chapman-Miller (1940) method relates to the later usage, which we recommended as preferable to the earlier usage. However, the *determinations* were correctly made in either case.

3. But by an oversight associated with staff changes in the computing bureau, then engaged on lunar geophysical computations under the author's direction, at the Imperial College, London, the working instructions for the calculation of the *probable errors* of the determinations made on the new (μ) basis remained for some time without the necessary simple changes of sign notified above. Thus the computed probable errors were erroneous. This applies to several of the determinations of the lunar atmospheric tide (in barometric pressure) published by Chapman and Tschu (1948); the *determinations* were correct, but not all the probable errors. The correct values of the probable errors will be given in a future paper by K. C. Westfold and myself.

4. This opportunity may be taken of explaining in simpler terms than has perhaps hitherto been done, the *nature* of the calculation of the probable errors. The determination, of which the worked-out details are given by Tschu (1949), will be taken as illustration; thus only the semidiurnal solar and lunar variations will be considered (the principles are, of course, the same in all cases). These daily variations are represented by

$$S_2 \sin (2t + \sigma_2) \text{ and } L_2 \sin (2\tau + \lambda_2)$$

respectively, where t , τ denote mean solar and mean lunar time, reckoned from lower (solar or lunar) transit, at the rate 2π per (solar or lunar) day; and

$$\tau = t - \nu,$$

where ν , a measure of the lunar phase, increases from 0 to 2π from one new (mean) moon to the next. The number μ referred to above is the nearest integer

to $24(1-\nu/2\pi)$, for the epoch of Greenwich mean noon on each Greenwich day. Thus, if we ignore the small variation in ν during one solar day (a neglect which is of course allowed for by a final correction in the *determination* of L_2 and λ_2), the average lunar semidiurnal variation on a day in the group with number μ is

$$L_2 \sin (2t + \lambda'_2),$$

where

$$\lambda'_2 = \lambda_2 - 2\nu = \lambda_2 + \mu\pi/6.$$

Thus the phase angle of this variation (regarded as solar semidiurnal) increases from each group to the next by $\pi/6$ or 30° .

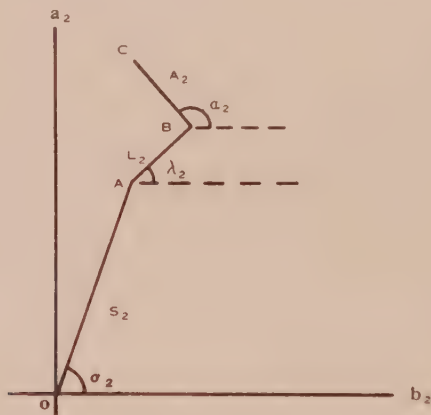


Fig. 1

If there were no accidental variations superposed on the regular solar and lunar semidiurnal variations, harmonic analysis for the group μ would give a solar semidiurnal term

$$C \sin (2t + c_2) = S_2 \sin (2t + \sigma_2) + L_2 \sin (2t + \lambda_2 + \mu\pi/6).$$

This may be represented on a harmonic dial as in Figure 1, where $C_2 \sin (2t + c_2)$ is represented by a vector OB of length C_2 and phase angle c_2 ; OA similarly represents the solar term $S_2 \sin (2t + \sigma_2)$, and AB the lunar term $L_2 \sin (2t + \lambda_2 + \mu\pi/6)$. For successive values of μ , 0 to 24, the points B in Figure 1 move counter-clockwise round A , in steps of 30° . Actually there will be an "accidental" superposed variation $A_2 \sin (2t + \alpha_2)$, represented by BC (where A_2 and α_2 cannot be foretold), so that the harmonics computed from the data will not be represented by OB but by OC . (In Figure 1, which is purely diagrammatic, OA is not drawn so long in comparison with AB as would correspond to the illustrative example in Tschu's (1949) paper.)

In the illustrative example, the numbers $A_{2\mu}$, N_μ and $B_{2\mu}$, N_μ in Tschu's Figure 6 represent the components along the a, b axes (see Fig. 1 of present communication) of the computed mean semidiurnal variation for the pairs of μ groups 0 and 12, 1 and 13, 2 and 14, . . . (the groups are taken in pairs because their phase angles λ'_2 are the same for both members of each pair): these a, b components are the projections of OC along the Oa_2 , Ob_2 axes respectively.

The mean values of these numbers $A_{2,r}/N_r$ and $B_{2,r}/N_r$ are the a_2, b_2 coordinates of the centroid A of the 12 points C for these 12 pairs of groups. The a_2, b_2 components of the vectors AC , representing $L_2 \sin (2t + \lambda'_2) + A_2 \sin (2t + \alpha_2)$ —



Fig. 2

where A_2, α_2 are unknown, and not the same for successive groups—correspond to the numbers $\Delta a_2, \Delta b_2$ in columns 3 and 4 of Tschu's Figure 6. These vectors AC are shown by their end-points C_0, C_1, C_2, \dots (corresponding to $\mu=0, 1, 2, \dots$)

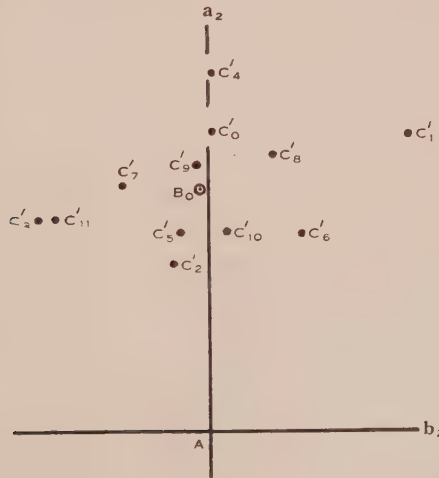


Fig. 3

in Figure 2. They show the expected counter-clockwise progression round A , but in a decidedly irregular way, owing to the accidental error.

In order to compute the probable error, the vectors AC_0, AC_1, AC_2, \dots are turned backward (clockwise) through $30\mu^\circ$, where $\mu=0, 1, 2, \dots$. This

gives the points C'_0, C'_1, C'_2, \dots in Figure 3 (C'_0 is of course identical with C_0). It should be noted that Figure 3 is on a somewhat larger scale than Figure 2: The distance AC_0 in Figure 2 corresponds to AC'_0 in Figure 3. These vectors AC' include $L_2 \sin(2t + \lambda_2)$, together with the reoriented vectors BC (now $B'C'$), which remain random after the systematic reorientation: it is assumed that the mean of the vectors $B'C'$ is zero, so that the centroid B_0 in Figure 3 refers to $L_2 \sin(2t - \lambda_2)$. The a_2, b_2 components AC' correspond to $\Delta a'_2, \Delta b'_2$ in Table 1. The vector AB_0 to the centroid of the points C'_0, C'_1, \dots is taken to represent $L_2 \sin(2t + \lambda_2)$ —except for appropriate final corrections.

The last column in Table 1 represents the lengths $B_0C'_0, B_0C'_1, \dots$, the accidental parts of the computed determination of L_2, λ_2 ; they correspond to the last column in Table 1, and may be determined either numerically, as there, or graphically. Their mean represents R' , in this case 0.88 units—much less than the mean distance L_2 of the points C' from A . If, however, they are large compared with L_2 , the change from Figure 2 to Figure 3 is not worth making, and the probable error can be determined from the mean distance of the points C_0, C_1, \dots in Figure 2 from A ; then Tschu's Table 3 may be used instead of completing the last five columns (as above) in Figure 6.

The errors of sign in Tschu's Table 4 had the effect of turning the vectors AC_0, AC_1, \dots through 30μ degrees in the counter-clockwise direction (which was the correct one when the lunar groups of days were ordered according to transit time instead of μ). This would give erroneous points C'_0, C'_1, \dots (represented by the incorrect 6th and 7th columns of Tschu's Fig. 6): as this incorrect reorientation does not bring the AB vectors into line, the mean centre of the erroneous points does not represent the lunar variation, which, however, had already been determined correctly, as shown in the left-hand portion of Figure 6. The incorrectly estimated probable error was enlarged because it included a good deal of what was really true lunar semidiurnal variation, oriented wrongly round A ; and this would always be so (save for exceptional distributions of the accidental error) when the lunar variation is not small compared with the probable error.

References

- BARTELS, J., and FANSELAU, G. (1938).—*Abh. Geophys. Inst. Potsdam*, No. 2. See also *Z. Geophys.* **13**: 311 (1937); *Terr. Magn. Atmos. Elect.* **43**: 155 (1938).
 CHAPMAN, S., and MILLER, J. C. P. (1940).—*Mon. Not. R. Astr. Soc. Geophys. Suppl.* **4**: 649.
 CHAPMAN, S., and TSCHU, K. K. (1948).—*Proc. Roy. Soc. A* **195**: 310.
 MILLER, J. C. P. (1934).—*Mon. Not. R. Astr. Soc.* **94**: 860.
 SCHMIDT, A. (1928).—*Veröff. preuss. met. Inst.* No. 357 (*Abh.* **9**: No. 1).
 TSCHU, K. K. (1949).—*Aust. J. Sci. Res. A* **2**: 1.

THE VELOCITY OF SOUND IN GASES AT LOW PRESSURES*

By R. L. ABBEY†

Recent measurements of the velocity of sound (Abbey and Barlow 1948; Maulard 1949), indicate that at pressures below 15 cm. Hg there is an increase of velocity unaccounted for by theory. In 1948 the author measured the velocity of sound in a tube 7 cm. in diameter at a frequency of 1000 c/s. A change in velocity at pressures below 20 cm. Hg was observed for several gases investigated. For air, the velocity increased by 9 m. sec.⁻¹ at a pressure of 0.5 cm. Hg. It was suggested that the increase might be due to an error in the tube correction.

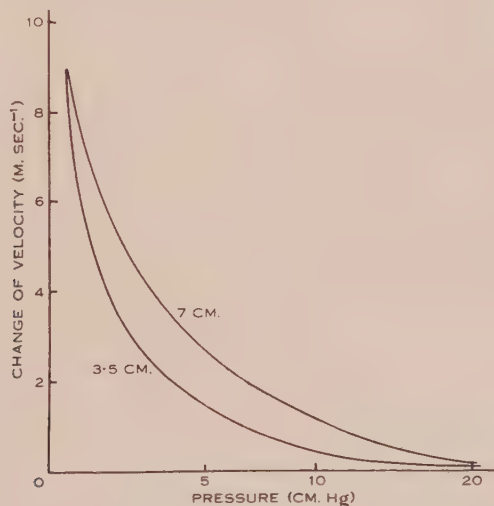


Fig. 1.—Velocity changes in 7 and 3.5 cm. diameter tubes.

Later, Maulard (1949), in France, measured the time taken for a pulse to travel a measured distance in a tube 8 cm. in diameter. The method was not sensitive enough for measurements to be made below a pressure of 4 cm. Hg, but at this pressure the velocity had increased by about 14 m. sec.⁻¹. The probable reason for this discrepancy between the results of Maulard and the author arises from the difference between the methods used, and will be discussed later in this communication.

As the tube correction is inversely proportional to both the diameter of the tube and the square root of the frequency, an error in this formula will be more appreciable for small diameters and low frequencies when the correction becomes large. Both these factors have been studied in a further series of measurements.

* Manuscript received October 24, 1951.

† Physics Department, University of Melbourne.

Effect of Tube Diameter

A new apparatus was constructed with the same principle of operation as previously, but with the following modifications in design. The tube diameter was reduced to 3.5 cm. It was not lagged but was painted black and measurements were made in a room where the temperature was constant. The tube did not have a rubber section as before, but the speaker was mounted inside the tube on three small pieces of sponge rubber. The microphone was a light crystal earphone, also mounted on sponge rubber. This construction gave a better

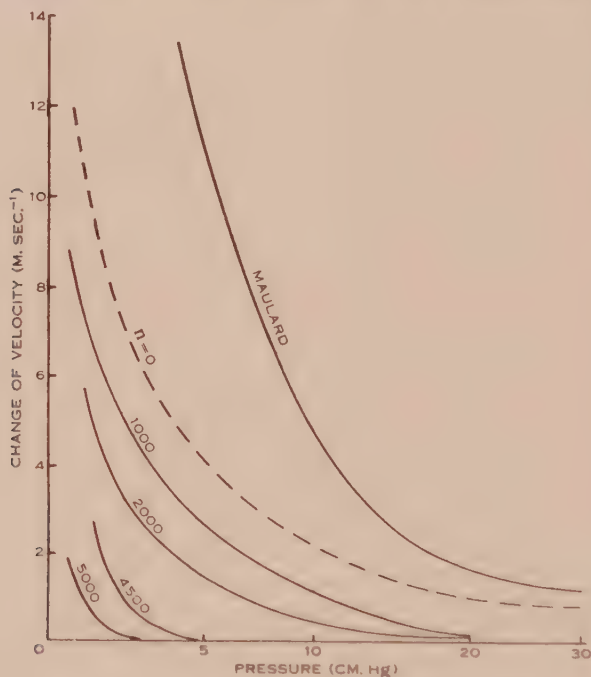


Fig. 2.—Velocity changes for different frequencies.

protection against the vibrations being transmitted through the walls of the tube. A microphone preamplifier increased the sensitivity for measurements at lower pressures.

The results of measurements in air are shown in Figure 1 compared with the previous results for the 7 cm. tube. The velocity increase is slightly less for the smaller diameter tube, but this is not considered significant within the limits of experimental error. These results indicate that the effect is not due to the tube correction alone.

Effect of Frequency

Using the apparatus with the 7 cm. tube described in the previous paper, the velocity of sound in air was measured at frequencies of 2000, 4500, and 5000 c/s. The different frequencies were obtained by changing the value of the inductance and capacitance in the tuned circuit of the amplifier. The results

of these measurements are shown in Figure 2. It was found that the velocity change becomes smaller as the frequency increases. Assuming this dependence on frequency continues uniformly for frequencies less than 1000 c/s., a curve for zero frequency was obtained. This limit is shown by the dotted curve in Figure 2.

This frequency dispersion gives a possible explanation for the higher increase of velocity with pressure reported by Maulard. His results, as shown in the curve in Figure 2 would correspond to a very low frequency. This agrees with his description of the pulse, the wavelength of which corresponds to a frequency of about 200 c/s. The observed increase of pulse length with distance of propagation would be expected both from the above effect and the tube correction which is also a function of the frequency. Maulard's results still lie beyond the zero frequency limit however, but experimental error (2 m. sec.^{-1}), the uncertainties associated with measuring a pulse containing a range of different velocities, and the possible error in extrapolation to zero frequency may account for this remaining discrepancy.

References

- ABBEY, R. L., and BARLOW, G. E. (1948).—*Aust. J. Sci. Res. A* **1**: 175.
MAULARD, J. (1949).—*C.R. Acad. Sci. Paris* **229**: 1, 25.

THE BENDING OF A SEMI-INFINITE STRIP

By R. C. T. SMITH*

[*Manuscript received September 24, 1951*]

Summary

An exact solution is obtained for the normal displacement of a thin semi-infinite strip subjected to arbitrary displacements and couples on the short edge but with the long edges clamped. It is expressed as a series in the characteristic functions of a matrix differential equation.

I. INTRODUCTION

The bending of a thin plate is governed by the equation

$$\nabla^4 w = \left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial t^2} \right)^2 w = 0, \quad \dots \quad (1)$$

for the case in which no normal pressure is applied to the plate (Love 1934). In (1), w is the normal deflection and y, t are rectangular coordinates in the surface. Here this equation is solved for the region

$$y > 0, \quad -1 < t < 1, \quad \dots \quad (2)$$

subject to the boundary conditions

$$w = \frac{\partial w}{\partial t} = 0, \quad \text{at } t = 1, \quad \dots \quad (3)$$

$$w = \frac{\partial w}{\partial t} = 0, \quad \text{at } t = -1, \quad \dots \quad (4)$$

(clamped-edge conditions) and

$$\frac{\partial^2 w}{\partial y^2} = f(t), \quad \frac{\partial^2 w}{\partial t^2} = g(t), \quad \text{at } y = 0, \quad \dots \quad (5)$$

where $f(t), g(t)$ are arbitrary independent functions of t determined, for example, if the displacements of the short edge and the couples about this edge are known. However to be consistent with (3), (4) we must have

$$\int_{-1}^1 g(t) dt = 0, \quad \dots \quad (5')$$

$$\int_{-1}^1 t g(t) dt = 0. \quad \dots \quad (5'')$$

* Department of Mathematics and Physics, New England University College, Armidale, N.S.W.

The boundary conditions suggest taking $\partial^2 w / \partial y^2$, $\partial^2 w / \partial t^2$ as unknowns. Separating the variables we write $\partial^2 w / \partial y^2$, $\partial^2 w / \partial t^2$ as the sum of particular solutions.

$$\frac{\partial^2 w}{\partial y^2} = \sum_s x_1(t) e^{-s y}, \quad \frac{\partial^2 w}{\partial t^2} = \sum_s x_2(t) e^{-s y}, \quad \dots \quad (6)$$

where $x_1(t)$, $x_2(t)$ depend also on s and the summations are to be taken over a set of values of s to be determined subsequently. Since

$$\frac{\partial^4 w}{\partial t^2 \partial y^2} - \frac{\partial^4 w}{\partial y^2 \partial t^2} = 0,$$

we have

$$\frac{d^2 x_1}{dt^2} - s^2 x_2 = 0.$$

From (1)

$$\frac{d^2 x_2}{dt^2} + s^2 (2x_2 + x_1) = 0.$$

Thus in matrix notation

$$\frac{d^2}{dt^2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + s^2 \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0. \quad \dots \quad (7)$$

From (3), (4)

$$x_1 = \frac{dx_1}{dt} = 0, \quad \text{at } t=1, \quad \dots \quad (3')$$

$$x_1 = \frac{dx_1}{dt} = 0, \quad \text{at } t=-1. \quad \dots \quad (4')$$

We will obtain an expansion formula of the type

$$\begin{bmatrix} f(t) \\ g(t) \end{bmatrix} = \sum_n c_n \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix}, \quad \dots \quad (8)$$

where $\{\varphi_1^{(n)}(t), \varphi_2^{(n)}(t)\}$ is an eigenvector of (7) with the boundary conditions (3'), (4') corresponding to the eigenvalue s_n^2 , and the c_n are constants. Then

$$\frac{\partial^2 w}{\partial y^2} = \sum c_n \varphi_1^{(n)}(t) e^{-s_n y}, \quad \frac{\partial^2 w}{\partial t^2} = \sum c_n \varphi_2^{(n)}(t) e^{-s_n y}$$

take the values (5) at $y=0$, and

$$w = \sum \frac{c_n}{s_n^2} \varphi_1^{(n)}(t) e^{-s_n y} \quad \dots \quad (9)$$

will satisfy the differential equation (1) and the boundary conditions (3), (4). The eigenvalue $s=0$ may need special attention.

Next we obtain a biorthogonality property which enables us to determine the c_n . We consider a generalized Wronskian defined by

$$W = \begin{vmatrix} y_1 & y_2 \end{vmatrix} \begin{bmatrix} dx_1/dt \\ dx_2/dt \end{bmatrix} - \begin{bmatrix} dy_1/dt & dy_2/dt \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \quad \dots \quad (10)$$

Obviously

$$\frac{dW}{dt} = [y_1 \ y_2] \begin{bmatrix} d^2x_1/dt^2 \\ d^2x_2/dt^2 \end{bmatrix} - [d^2y_1/dt^2 \ d^2y_2/dt^2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \quad \dots \quad (11)$$

If $\{x_1, x_2\}$ is a solution of (7), (11) becomes

$$\frac{dW}{dt} = - \left\{ s^2 [y_1 \ y_2] \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} + [d^2y_1/dt^2 \ d^2y_2/dt^2] \right\} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

We therefore define the adjoint equation to (7) as

$$\frac{d^2}{dt^2} [y_1 \ y_2] + s^2 [y_1 \ y_2] \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} = 0. \quad \dots \quad (12)$$

If W is to vanish automatically whenever $x_1 = dx_1/dt = 0$ we must also have $y_2 = dy_2/dt = 0$. The boundary conditions adjoint to (3'), (4') are therefore

$$y_2 = \frac{dy_2}{dt} = 0, \quad \text{at } t=1, \quad \dots \quad (13)$$

$$y_2 = \frac{dy_2}{dt} = 0, \quad \text{at } t=-1. \quad \dots \quad (14)$$

Now if $[y_1 \ y_2] = [\psi_1^{(m)}(t) \ \psi_2^{(m)}(t)]$ is an eigenvector of (12) with the boundary conditions (13), (14) corresponding to s_m^2 and

$$\{x_1, x_2\} = \{\varphi_1^{(n)}(t), \varphi_2^{(n)}(t)\}$$

is an eigenvector of (7) with the boundary conditions (3'), (4') corresponding to s_n^2 , we have

$$(s_m^2 - s_n^2) [\psi_1^{(m)}(t) \ \psi_2^{(m)}(t)] \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix} = \frac{dW}{dt}.$$

Integrating and using the boundary conditions

$$(s_m^2 - s_n^2) \int_{-1}^1 [\psi_1^{(m)}(t) \ \psi_2^{(m)}(t)] \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix} dt = 0. \quad \dots \quad (15)$$

Thus if $s_m^2 \neq s_n^2$

$$\int_{-1}^1 [\psi_1^{(m)}(t) \ \psi_2^{(m)}(t)] \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix} dt = 0. \quad \dots \quad (16)$$

Unpublished work of the author's shows that any eigenvalue s_n^2 of the system (7), (3'), (4') is also an eigenvector of the adjoint system (12), (13), (14) and also that both systems have the same number of eigenvectors for any particular s_n^2 . If each of the systems (7), (3'), (4'), and (12), (13), (14) has more than one eigenvector corresponding to a particular s_n^2 , it is possible in general to choose the eigenvectors so that suitably numbered pairs satisfy the biorthogonality condition (16) for m, n different, that is, for

$$[\psi_1^{(m)}(t) \ \psi_2^{(m)}(t)], \{\varphi_1^{(n)}(t), \varphi_2^{(n)}(t)\},$$

from different pairs. If we assume $\{f(t), g(t)\}$ can be expanded in a series of the form

$$\begin{bmatrix} f(t) \\ g(t) \end{bmatrix} = \sum c_n \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix}, \quad \dots \quad (8)$$

and that term by term integration can be justified we obtain a formula for the coefficients c_n

$$c_n = \frac{1}{k_n} \int_{-1}^1 [\psi_1^{(n)}(t) \quad \psi_2^{(n)}(t)] \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} f(t) \\ g(t) \end{bmatrix} dt, \dots \quad (17)$$

where

$$k_n = \int_{-1}^1 [\psi_1^{(n)}(t) \quad \psi_2^{(n)}(t)] \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix} dt, \dots \quad (18)$$

It is quite straightforward to determine the eigenvectors and only the results will be given. Because of symmetry we see that if $\{\varphi_1(t), \varphi_2(t)\}$ is an eigenvector, then so is $\{\varphi_1(-t), \varphi_2(-t)\}$. Consequently

$\frac{1}{2}\{\varphi_1(t), \varphi_2(t)\} + \frac{1}{2}\{\varphi_1(-t), \varphi_2(-t)\}$ and $\frac{1}{2}\{\varphi_1(t), \varphi_2(t)\} - \frac{1}{2}\{\varphi_1(-t), \varphi_2(-t)\}$ are even and odd eigenvectors respectively. We need only consider even and odd eigenvectors.

II. EVEN EIGENFUNCTIONS

(i) $s=0$

$$\begin{aligned} \varphi_1 &= 0, & \varphi_2 &= 1, \\ \psi_1 &= 1, & \psi_2 &= 0; \\ k &= -2. \end{aligned}$$

This eigenfunction is of no interest in the present connection since from (5) the coefficient corresponding to it vanishes.

(ii) $\sin 2s = -2s, (s \neq 0)$

$$\begin{aligned} \varphi_1 &= s \sin s \cos st - st \cos s \sin st, \\ \varphi_2 &= -(s \sin s + 2 \cos s) \cos st + st \cos s \sin st, \\ \psi_1 &= (s \sin s - 2 \cos s) \cos st - st \cos s \sin st, \\ \psi_2 &= s \sin s \cos st - st \cos s \sin st, \\ k &= -4 \cos^4 s. \end{aligned}$$

III. ODD EIGENFUNCTIONS

(i) $s=0$

$$\begin{aligned} \varphi_1 &= 0, & \varphi_2 &= t, \\ \psi_1 &= t, & \psi_2 &= 0, \\ k &= -\frac{2}{3} \end{aligned}$$

This eigenvector is of no interest in the present connection since from (5') the coefficient corresponding to it vanishes.

(ii) $\sin 2s = 2s, (s \neq 0)$

$$\begin{aligned} \varphi_1 &= s \cos s \sin st - st \sin s \cos st, \\ \varphi_2 &= -(s \cos s - 2 \sin s) \sin st + st \sin s \cos st, \\ \psi_1 &= (s \cos s + 2 \sin s) \sin st - st \sin s \cos st, \\ \psi_2 &= s \cos s \sin st - st \sin s \cos st, \\ k &= -4 \sin^4 s. \end{aligned}$$

From a practical point of view the only further information needed is on the distribution of the eigenvalues. This is reasonably complete. Hillman and Salzer (1943) have published a table of the first 10 roots of $\sin z = z$. For large $|z|$ the n th root with positive real and imaginary parts is given approximately by

$$(2n+0.5)\pi + i \log (4n+1)\pi.$$

Robbins and Smith (1948) have published a similar table of roots of $\sin z = -z$. For large $|z|$ the n th root with positive real and imaginary parts is given approximately by

$$(2n+1.5)\pi + i \log (4n+3)\pi.$$

In each case the roots occur in conjugate pairs.

A complete formal solution of the problem stated in equations (1) to (5) is given by (9), (17), (18), and the expressions for the eigenvectors. This is obviously (and trivially) valid if the sum in (8) contains only a finite number of terms but, in general, for arbitrary $\{f(t), g(t)\}$ there is no reason to expect the series (8) to converge. In fact, for large n ,

$$\varphi_1 = O(n^{3/2}e^{vt}) = O(n^2),$$

$$\varphi_2 \approx -\varphi_1, \quad \psi_1 \approx \varphi_1, \quad \psi_2 = \varphi_1, \quad k \approx 4n^2\pi^2,$$

where v is the imaginary part of s_n . Assuming bounded variation of $f(t), g(t)$ we can obtain $c_n = O(n^{-1})$ consequently $c_n \varphi_1^{(n)} = O(n)$. Thus the series $\sum c_n \varphi_1^{(n)}(t), \sum c_n \varphi_2^{(n)}(t)$ may well diverge unless $f(t), g(t)$ satisfy stronger conditions. This does not necessarily affect the practical value of the solution since the series

$$\sum c_n \varphi_1^{(n)}(t) e^{-s_n y}, \quad \sum c_n \varphi_2^{(n)}(t) e^{-s_n y}, \quad \sum \frac{c_n}{s_n} \varphi_1^{(n)}(t) e^{-s_n y},$$

will converge rapidly for any $y > 0$ and we might expect a summability relation of the type

$$\lim_{y \rightarrow 0} \sum c_n \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix} e^{-s_n y} = \begin{bmatrix} f(t) \\ g(t) \end{bmatrix}.$$

If convergence is uniform with respect to y we may take $y = 0$ and obtain

$$\sum c_n \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix} = \begin{bmatrix} f(t) \\ g(t) \end{bmatrix}. \quad \dots\dots\dots (8)$$

Appendix I establishes (8) under certain conditions.

IV. ACKNOWLEDGMENT

The author wishes to thank Mr. L. M. Sheppard for checking the tedious matrix multiplications involved in Appendix I.

V. REFERENCES

- HILLMAN, A. P., and SALZER, H. E. (1943).—Roots of $\sin z = z$. *Phil. Mag.* (7) **34**: 575.
 LOVE, A. E. H. (1934).—“The Mathematical Theory of Elasticity.” p. 487 et seq. (Cambridge Univ. Press.)

ROBBINS, C. I., and SMITH, R. C. T. (1948).—A table of roots of $\sin z = -z$. *Phil. Mag.* (7) **39**: 1005.

TITCHMARSH, E. C. (1946).—“Eigenfunction Expansions Associated with Second Order Differential Equations.” (Oxford Univ. Press.)

APPENDIX I

The expansion formula (8) will be justified for somewhat restrictive conditions on $f(t)$, $g(t)$, namely that

$$f(-1) - g(-1) - f'(-1) - g'(-1) = 0, \quad f(1) - g(1) - f'(1) - g'(1) = 0$$

and that $f''(t)$, $g''(t)$ are of bounded variation. Unpublished work of the author's on matrix differential equations develops formulae analogous to those used by Titchmarsh (1946). By variation of parameters it can be shown that the solution of the non-homogeneous equation

$$\frac{d^2 M}{dt^2} + s^2 \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} M = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} f(t) \\ g(t) \end{bmatrix} = K(t), \text{ say,} \quad \dots \quad (\text{i})$$

subject to the boundary conditions

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} M + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \frac{dM}{dt} = 0, \quad \begin{cases} \text{at } t=1, & \dots \dots \dots (\text{ii}) \\ \text{at } t=-1, & \dots \dots \dots (\text{iii}) \end{cases}$$

is given by

$$M = X_1(t) W_{1,2}^{-1} \int_{-1}^t Y_2(u) K(u) du + X_2(t) W_{2,1}^{-1} \int_t^1 Y_1(u) K(u) du, \quad \dots \dots \dots (\text{iv})$$

where the 2×2 matrices $X_1(t)$, $X_2(t)$ satisfy the homogeneous equation

$$\frac{d^2 X}{dt^2} + s^2 \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} X = 0,$$

and $X_1(t)$ the boundary conditions (ii), $X_2(t)$ the boundary condition (iii); the 2×2 matrices $Y_1(t)$, $Y_2(t)$ satisfy the adjoint equation

$$\frac{d^2 Y}{dt^2} + s^2 Y \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} = 0,$$

and the adjoint boundary conditions

$$Y \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \frac{dY}{dt} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 0, \quad \begin{cases} Y_1 \text{ at } t=1, \\ Y_2 \text{ at } t=-1. \end{cases}$$

$$W_{1,2} = Y_2 \frac{dX_1}{dt} - \frac{dY_2}{dt} X_1,$$

$$W_{2,1} = Y_1 \frac{dX_2}{dt} - \frac{dY_1}{dt} X_2.$$

We now give the more important properties of the matrix Wronskian

$$W = Y \frac{dX}{dt} - \frac{dY}{dt} X.$$

where X, Y are square matrices. If X satisfies a boundary condition of the Sturm-Liouville type at a given point

$$CX + S \frac{dX}{dt} = 0,$$

the vanishing of the Wronskian is equivalent to the condition that Y satisfies an adjoint boundary condition of the form

$$YC^* + \frac{dY}{dt} S^* = 0,$$

provided the rank of $\left[X, \frac{dX}{dt} \right]$ is equal to the order of X . Thus

$$W_{1,1} = Y_1 \frac{dX_1}{dt} - \frac{dY_1}{dt} X_1,$$

vanishes at $t=1$ and

$$W_{2,2} = Y_2 \frac{dX_2}{dt} - \frac{dY_2}{dt} X_2,$$

vanishes at $t=-1$. Similarly if Y satisfies the adjoint boundary condition and W vanishes, X satisfies the original boundary condition (provided the rank of $\left[Y, \frac{dY}{dt} \right]$ is equal to the order of Y). More generally, if the rank of W is less than its order, X satisfies the given boundary condition, and the above condition on the rank of $\left[X, \frac{dX}{dt} \right]$ is satisfied, then a linear combination of the rows of Y satisfies the adjoint boundary condition. There is a similar result interchanging the roles of X and Y . Thus if $W_{2,1}$ is singular at $t=1$ a linear combination of the columns of X_2 satisfies the boundary condition (ii). Since X_2 satisfies (iii) the linear combination found satisfies both (ii) and (iii), i.e. it is an eigenvector. So far we have not connected the Wronskian with the differential equation. If we assume Y satisfies an equation of the form

$$\frac{d^2 X}{dt^2} + AX = 0,$$

and Y the adjoint equation

$$\frac{d^2 Y}{dt^2} + YA = 0,$$

then we easily see that

$$\frac{dW}{dt} = 0,$$

that is, W is constant. Thus all the Wronskians $W_{1,1}=0$, $W_{1,2}$, $W_{2,1}$, $W_{2,2}=0$ are constant matrices. Summing up if $W_{2,1}$ is singular there is at least one eigenvector of the system (7), (3'), (4'), and the same number of eigenvectors of the adjoint system (12), (13), (14). From the existence of the eigenvectors it can be easily shown that $W_{1,2}$ is also singular.

For large $|s|$ we may expect $\frac{d^2 M}{dt^2}$ to be negligible compared with the other term on the left-hand side of (i), that is,

$$M \approx \frac{1}{s^2} \begin{bmatrix} f(t) \\ g(t) \end{bmatrix}. \quad \dots\dots\dots (\text{v})$$

Thus for a sequence of suitably chosen contours D_N whose minimum distance from the origin tends to infinity we may hope for a relation

$$\lim_{N \rightarrow \infty} \frac{1}{2\pi i} \int_{D_N} Ms \, ds = \begin{bmatrix} f(t) \\ g(t) \end{bmatrix}. \quad \dots\dots\dots (\text{vi})$$

For the problem we are considering it is easy to see that M is a meromorphic function of s with the zeros of $\det W_{1,2}$, $\det W_{2,1}$, as poles. The contours D_N must avoid these poles. Now the integral (vi) can also be evaluated by residues. The general theory shows that the zeros of $\det W_{1,2}$ coincide with those of $\det W_{2,1}$ and are, in fact, the eigenvalues of the system (7), (3'), (4'); further that the residue of a pole s_n (and at the pole $-s_n$) is simply

$$\frac{1}{2} c_n \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix}, \quad \dots\dots\dots (\text{vii})$$

where c_n is given by (17). If the contour D_N encloses the first $2N$ poles s_n ($s=0$ being counted twice), with non-negative real parts and the corresponding $(2N-2)$ poles $-s_n$, then

$$\frac{1}{2\pi i} \int_{D_N} Ms \, ds = \sum_1^{2N} c_n \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix}. \quad \dots\dots\dots (\text{viii})$$

(This relation is quite straightforward, though tedious, to verify by explicit calculation. The calculation is simplified by postmultiplying X_1 by a suitable matrix so that the first column of the product is an eigenvector for $s=s_n$ and premultiplying Y_2 by a matrix such that the first row of the product is also an eigenvector. The second term in (iv) is treated similarly.) Thus if we can justify (vi) for $[f(t), g(t)]$ satisfying suitable conditions, we have shown that

$$\begin{bmatrix} f(t) \\ g(t) \end{bmatrix} = \lim_{N \rightarrow \infty} \sum_1^{2N} c_n \begin{bmatrix} \varphi_1^{(n)}(t) \\ \varphi_2^{(n)}(t) \end{bmatrix}. \quad \dots\dots\dots (\text{ix})$$

This is, of course, (8).

We will now investigate (vi) in detail. We take

$$X_1 = X(t-1), \quad X_2 = X(t+1),$$

where

$$X(t) = \begin{bmatrix} \sin st - st \cos st & st \sin st \\ \sin st + st \cos st & 2 \cos st - st \sin st \end{bmatrix},$$

$$Y_1 = Y(t-1), \quad Y_2 = Y(t+1),$$

where

$$Y(t) = \begin{bmatrix} 2 \cos st + st \sin st & st \sin st \\ 3 \sin st - st \cos st & \sin st - st \cos st \end{bmatrix}.$$

It can be easily verified that $W_{1,2}$, $W_{2,1}$ are independent of t . Taking $t=0$, we obtain

$$\begin{aligned} W_{1,2} &= Y(1) \frac{dX(-1)}{dt} - \frac{dY(1)}{dt} X(-1), \\ &= 2s \begin{bmatrix} 2s \sin 2s & -\sin 2s - 2s \cos 2s \\ \sin 2s - 2s \cos 2s & -2s \sin 2s \end{bmatrix}, \\ W_{2,1} &= Y(-1) \frac{dX(1)}{dt} - \frac{dY(-1)}{dt} X(1), \\ &= 2s \begin{bmatrix} 2s \sin 2s & \sin 2s + 2s \cos 2s \\ -\sin 2s + 2s \cos 2s & -2s \sin 2s \end{bmatrix}, \\ W_{1,2}^{-1} &= \frac{1}{2s\{\sin^2 2s - 4s^2\}} \begin{bmatrix} -2s \sin 2s & \sin 2s + 2s \cos 2s \\ -\sin 2s + 2s \cos 2s & 2s \sin 2s \end{bmatrix}, \\ W_{2,1}^{-1} &= \frac{1}{2s\{\sin^2 2s - 4s^2\}} \begin{bmatrix} -2s \sin 2s & -\sin 2s - 2s \cos 2s \\ \sin 2s - 2s \cos 2s & 2s \sin 2s \end{bmatrix}. \end{aligned}$$

The probable asymptotic behaviour of M suggests that we take

$$M = \frac{1}{s^2} \begin{bmatrix} f(t) \\ g(t) \end{bmatrix} + P, \quad \dots\dots\dots (\mathbf{x})$$

Since we assume that

$$f(-1)=f(1)=g(-1)=g(1)=0 \text{ and } f'(-1)=f'(1)=g'(-1)=g'(1)=0,$$

P satisfies the same boundary conditions as M . Also from the differential equation (i)

$$\frac{d^2 P}{dt^2} + s^2 \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} P = -\frac{1}{s^2} \begin{bmatrix} f''(t) \\ g''(t) \end{bmatrix}.$$

The solution of this equation is given by (iv).

Finally we obtain

$$\begin{aligned} Ms &= \frac{1}{s} \begin{bmatrix} f(t) \\ g(t) \end{bmatrix} - \frac{1}{s} X_1(t) W_{1,2}^{-1} \int_{-1}^t Y_2(u) \begin{bmatrix} f''(u) \\ g''(u) \end{bmatrix} du + \frac{1}{s} X_2(t) W_{2,1}^{-1} \int_t^1 Y_1(u) \begin{bmatrix} f''(u) \\ g''(u) \end{bmatrix} du. \\ &\dots\dots\dots (\mathbf{xi}) \end{aligned}$$

Thus to justify (vi) we have only to show that the integrals of the second and third terms on the right-hand side of (xi) round the contour D_N tend to zero as $N \rightarrow \infty$. As the contour D_N we take the square with vertices $2N\pi(\pm 1 \pm i)$. The contour is described in the anti-clockwise sense. For definiteness, we will consider

$$\frac{1}{s} X_1(t) W_{1,2}^{-1} \int_{-1}^t Y_2(u) \begin{bmatrix} f''(u) \\ g''(u) \end{bmatrix} du. \quad \dots\dots\dots (\mathbf{xii})$$

Expanding (xii) we obtain

$$\begin{aligned} & \frac{s(t-1)}{\sin^2 2s - 4s^2} \begin{bmatrix} -\cos s(t-1) & \sin s(t-1) \\ \cos s(t-1) & -\sin s(t-1) \end{bmatrix} \begin{bmatrix} \sin 2s & \cos 2s \\ \cos 2s & \sin 2s \end{bmatrix} \\ & \int_{-1}^t \begin{bmatrix} -\sin s(u-1) & \sin s(u-1) \\ \cos s(u+1) & \cos s(u+1) \end{bmatrix} \begin{bmatrix} (u-1)f''(u) \\ (u+1)g''(u) \end{bmatrix} du \\ & + \dots \text{(smaller terms).} \end{aligned}$$

The leading term in (xii) reduces to

$$\frac{s(t-1)}{\sin^2 2s - 4s^2} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \int_{-1}^t \cos s(t-u) \begin{bmatrix} (u+1)f''(u) \\ (u-1)g''(u) \end{bmatrix} du. \quad \dots\dots\dots \text{(xiii)}$$

Now consider

$$\int_{-1}^t \cos s(u+1)(u+1)f''(u)du.$$

Assuming $f''(u)$ to be of bounded variation we can write

$$(u+1)f''(u) = \varphi(u) - \psi(u),$$

where $\varphi(u)$, $\psi(u)$ are positive, non-decreasing functions of t .

The second mean value theorem shows that

$$\begin{aligned} \int_{-1}^t \mathbf{R}[\cos s(t-u)]\varphi(u)du &= \varphi(t) \int_{u^*}^t \mathbf{R}[\cos s(t-u)]du, \quad -1 < u^* < t, \\ &= \varphi(t) \mathbf{R} \left[\frac{1}{s} \sin s(t-u^*) \right] \\ &= O \left(\frac{e^{\tau(t+1)} + e^{-\tau(t+1)}}{|s|} \right). \end{aligned}$$

The same argument applies to the imaginary part of

$$\int_{-1}^t \cos s(t-u)\varphi(u)du$$

and we finally conclude that (xiii) is of order

$$\frac{e^{\tau(t+1)} + e^{-\tau(t+1)}}{|\sin^2 2s - 4s^2|}.$$

On D_N we easily see that

$$\frac{e^{\tau(t+1)} + e^{-\tau(t+1)}}{|\sin^2 2s - 4s^2|} = o \left(\frac{1}{s} \right).$$

The same result of course holds for $\psi(u)$. With the same conditions on $g''(u)$ as on $f''(u)$, we finally conclude that (xiii) is $o \left(\frac{1}{s} \right)$. Hence the integral of the leading term in (xii) round the contour D_N tends to zero as $N \rightarrow \infty$. The remaining terms in (xii) can be dealt with even more simply and likewise make

zero contribution (in the limit) to $\int_{D_N} Ms \, ds$. Thus (vi) has been proved under the conditions stated above and we have proved the convergence of the expansion (8).

Details of the calculation make it unlikely that the conditions imposed on $[f(t), g(t)]$ can be much relaxed. To obtain (x) we assume that $[f(t), g(t)]$, $[f'(t), g'(t)]$ are continuous.

If

$$[f(t-0), g(t-0)] \neq [f(t+0), g(t+0)],$$

or

$$[f'(t-0), g'(t-0)] \neq [f'(t+0), g'(t+0)],$$

additional terms of order

$$O\left(\frac{s \sin 2s \sin 2st}{\sin^2 2s - 4s^2}\right) \text{ or } O\left(\frac{\cos 2s \sin 2st}{\sin^2 2s - 4s^2}\right)$$

appear. These are not of $o\left(\frac{1}{s}\right)$ everywhere on D_N .

CURVE-FITTING BY THE METHOD OF GROUPING

By P. G. GUEST*

[*Manuscript received August 22, 1951*]

Summary

A method of fitting polynomials is described in which the "normal" equations are obtained much more rapidly than the corresponding equations in the least-squares method. Efficiencies are found to be about 90 per cent. The method is illustrated by an example.

I. INTRODUCTION

The method of curve-fitting developed in the present paper is really an extension of the method of grouping described by Jeffreys (1948, p. 193) for the estimation of the slope of a straight line. The saving of time in comparison with the standard least-squares method comes about in two ways. Firstly, in forming the "normal" equations the moments $\Sigma y(x)x^j$ are replaced by quantities obtained by simple addition and subtraction of the observations $y(x)$; and secondly, in fitting a polynomial of degree p only the first p powers of x instead of the first $2p$ powers are required. Provided the observations are spaced reasonably uniformly throughout the range of x covered, the efficiencies of the coefficients and fitted values are about 90 per cent.

A method of fitting straight lines and parabolas which has some resemblance to the present method has been described by Nair and Shrivastava (1942). In their method the "normal" equations are formed by grouping the observations into $p+1$ separate groups, with no observation appearing in more than one group. Although this last restriction simplifies the formation of the "normal" equations, the efficiencies of the estimates are greatly reduced. For the second degree coefficient the figure is about 60 per cent.

The method described here is intended to be used in the general case when the interval between successive observations is not constant. If the observations are equally spaced, it becomes possible to tabulate weight-functions for different values of n , the number of observations. The procedure using tabulated weight-functions in the equally-spaced case, which I have called "the method of weighted grouping" (Guest 1951), is more rapid than that using the "normal" equations which is described here, just as the procedure using tables of orthogonal polynomials is more rapid in the least-squares method.

* Physics Department, University of Sydney.

II. ESTIMATION OF THE POLYNOMIAL COEFFICIENTS

To form an unbiased estimate of the coefficient of x^j in the polynomial of degree p in x which is to fit the observed values $y(x)$, each observation must be multiplied (directly or indirectly) by a weight $w_{pj}(x)$ so chosen that

$$\sum w_{pj}(x)x^k = 0, \quad k \leq p, \quad k \neq j, \quad \dots \dots \dots (1)$$

the sum being taken over the n observations. The estimate of the coefficient of x^j is then

$$b_{pj} = \frac{\sum w_{pj}(x)y(x)}{\sum w_{pj}(x)x^j}, \quad \dots \dots \dots (2)$$

and the fitted polynomial is

$$u_p(x) = \sum_{j=0}^p b_{pj}x^j. \quad \dots \dots \dots (3)$$

Suppose we take a set of $p+1$ arbitrary functions $W_j(x)$ (no two of which are identical at all points of observation), and form the sums $\sum W_j(x)u_p(x)$ and $\sum W_j(x)y(x)$. If we equate these sums, we obtain a set of $p+1$ equations in the $p+1$ unknowns b_{pj} . We shall call these equations the "normal" equations. It is easy to show that the values obtained by solving these equations are unbiased estimates of the coefficients. For

$$b_{pj} = \frac{\sum \Delta_{kj} \sum W_k(x)y(x)}{\sum \Delta_{kj} \sum W_k(x)x^j}, \quad \dots \dots \dots (4)$$

where Δ_{kj} is the cofactor of $\sum W_k x^j$ in the determinant $|\sum W_k x^j|$. Hence

$$b_{pj} = \frac{\sum w_{pj}(x)y(x)}{\sum w_{pj}(x)x^j}, \quad \dots \dots \dots (5)$$

where

$$w_{pj}(x) = \sum_k \Delta_{kj} W_k(x). \quad \dots \dots \dots (6)$$

Also

$$\begin{aligned} \sum_x w_{pj}(x)x^m &= \sum_k \Delta_{kj} \sum_x W_k(x)x^m, \quad \dots \dots \dots (7) \\ &= 0, \quad \text{if } j \neq m, \end{aligned}$$

since the right-hand side of (7) is then the expansion of a determinant in terms of alien cofactors.

In the least-squares method, the weight-functions $W_j(x)$ are usually the powers x^j or the orthogonal polynomials $T_j(x)$. It can be shown that the least-squares method is the most efficient of all possible methods—that is, it gives the smallest standard errors.

In the method of grouping, the observations in certain ranges of x are combined into groups, and these groups added, omitted, or subtracted, no multiplications being involved. The weight-functions in this case are simply step-functions; $W_j(x)$ takes the values $+1$, 0 , and -1 . By analogy with the orthogonal polynomial $T_j(x)$ (which has j zeros) and the finite difference formula

for the coefficient of degree j (which contains $j+1$ terms), it would appear that $j+1$ groups (not counting those with zero weight) are required in the j th degree equation to obtain reasonable efficiencies. We will adopt this scheme of grouping.

III. CHOICE OF GROUPS

It is clear that the optimum method of grouping will depend on the particular values x at the points of observation, and hence may be different in different examples with the same value of n . The procedure adopted here is to determine the optimum method of grouping for the special case when the observations are equally spaced, and use the same method in the general case when the spacing is not uniform.

In this section, then, we shall assume that the observations are equally spaced. To simplify the calculations we shall suppose the independent variable x to be chosen so that its origin lies at the centre of the range and so that the interval between successive observations is unity. The values of x at the points of observation are then the integers or half-integers $+\frac{1}{2}(n-1)$ to $-\frac{1}{2}(n-1)$. The obvious procedure is to make $W_j(x)$ symmetrical, that is, to take $W_j(-x) = (-1)^j W_j(x)$.

In evaluating the efficiencies for various methods of grouping, we require formulae for the sums of powers at integral or half-integral values of x . These are collected below:

$$\begin{aligned} \sum_{x=-\frac{a-1}{2}}^{\frac{a-1}{2}} x^0 &= a = 2n, \text{ say,} \\ \sum_{x=-\frac{a-1}{2}}^{\frac{a-1}{2}} x^2 &= \frac{a^2-1}{12} = \frac{n^2-1}{4} x^2, \\ \sum_{x=-\frac{a-1}{2}}^{\frac{a-1}{2}} x^4 &= \frac{a(a^2-1)}{12} = \frac{n(n^2-1)}{4} x^4, \\ \sum_{x=-\frac{a-1}{2}}^{\frac{a-1}{2}} x^6 &= \frac{a(a^2-1)(3a^2-7)}{240} = \frac{n(n^2-1)(3n^2-7)}{240} x^6. \end{aligned}$$

(a) Polynomial of the First Degree

The equation determining b_{11} is

$$\Sigma W_1 y = b_{11} \Sigma W_1 x, \quad \dots \dots \dots (8)$$

and so

$$\frac{\sigma(y)}{\sigma(b_{11})} = \frac{\Sigma W_1 x}{\{\Sigma W_1^2\}^{\frac{1}{2}}}, \quad \dots \dots \dots (9)$$

Now

$$\Sigma W_1 f(x) = \left\{ \sum_{0, \frac{1}{2}}^{\frac{n-1}{2}} - \sum_{0, \frac{1}{2}}^{\frac{a-1}{2}} \right\} \{f(+x) - f(-x)\},$$

a being an integral parameter determining the location of the group. Setting $a = \alpha n$,

$$\Sigma W_1 x = \frac{n^2 - 1}{4} (1 - \alpha^2),$$

$$\Sigma W_1^2 = n(1 - \alpha),$$

and

$$\frac{\sigma(y)}{\sigma(b_{11})} = \frac{(n^2 - 1)(1 - \alpha^2)}{4\sqrt{n(1 - \alpha)}}.$$

The function on the right is a maximum when $\alpha = 1/3$. The number n_1 of observations in the group is then $\frac{1}{2}(n-1) - \frac{1}{2}(a-1) = n/3$. That is, in forming the first degree equation, we take the first $n_1 = n/3$ observations, omit the central $v_1 = n/3$ observations, and subtract the last n_1 observations.

If we denote the least-squares coefficient by a prime superscript, then

$$\left[\frac{\sigma(y)}{\sigma(b'_{11})} \right]^2 = \frac{n(n^2 - 1)}{12},$$

and so we have for the efficiency $\eta(b_{11})$ of the grouping estimate

$$\eta(b_{11}) = \frac{12(1 - \alpha^2)^2}{16(1 - \alpha)} = \frac{8}{9}.$$

The equation for b_{10} is $b_{10} = \Sigma y/n$, and $\eta(b_{10})$ is unity. These results are well known (Jeffreys 1948, p. 193).

(b) *Polynomial of the Second Degree*

The equations determining b_{20} , b_{22} are

$$\left. \begin{aligned} b_{20} \Sigma W_0 + b_{22} \Sigma W_0 x^2 - \Sigma W_0 y, \\ b_{20} \Sigma W_2 + b_{22} \Sigma W_2 x^2 - \Sigma W_2 y. \end{aligned} \right\} \dots\dots\dots (10)$$

The equation determining b_{21} is the same as the equation for b_{11} in the case $p=1$, and the function W_1 will be that found in Section III (a).

It is an advantage if W_0 can be permitted to retain the same form as for the cases $p=0$, $p=1$, and we accordingly set $W_0=1$ for all x . The standard errors of the coefficients are given by the equations

$$\frac{\sigma(y)}{\sigma(b_{22})} = \frac{\Sigma W_0 \Sigma W_2 x^2 - \Sigma W_0 x^2 \Sigma W_2}{[\Sigma W_2^2 (\Sigma W_0)^2 - \Sigma W_0^2 (\Sigma W_2)^2]^{\frac{1}{2}}}, \dots\dots\dots (11)$$

$$\frac{\sigma(y)}{\sigma(b_{20})} = \frac{\Sigma W_0 \Sigma W_2 x^2 - \Sigma W_0 x^2 \Sigma W_2}{[(\Sigma W_2 x^2)^2 \Sigma W_0^2 + (\Sigma W_0 x^2)^2 \Sigma W_2^2 - 2 \Sigma W_2 \Sigma W_0 x^2 \Sigma W_2 x^2]^{\frac{1}{2}}}, \dots\dots\dots (12)$$

$\Sigma W_2 f(x)$ will be of the form

$$\left\{ \sum_{\frac{1}{2}}^{\frac{n-1}{2}} - \sum_{\frac{1}{2}}^{\frac{a-1}{2}} - \sum_{\frac{1}{2}}^{\frac{b-1}{2}} \right\} f(x).$$

For example

$$\Sigma W_2 x^2 \doteq \frac{n(n^2-1)}{12} (1-\alpha^3-\beta^3),$$

where $\alpha=a/n$, $\beta=b/n$. Then

$$\frac{\sigma(y)}{\sigma(b_{22})} \doteq \frac{(n^2-1)\sqrt{n}}{12} \frac{(1-\alpha^3-\beta^3)-(1-\alpha-\beta)}{[(1-\alpha+\beta)-(1-\alpha-\beta)^2]^{\frac{1}{2}}},$$

where the parameters α , β are to be determined. This expression is a complicated function of α and β , and the location of its maximum is done most easily by successive approximations. A preliminary investigation shows that the maximum lies near $\alpha=0.78$, $\beta=0.51$. By evaluating the function at $(0.79, 0.51)$, $(0.77, 0.51)$, etc. we can find $\frac{\partial f}{\partial \alpha}$, $\frac{\partial f}{\partial \beta}$, $\frac{\partial}{\partial \alpha} \left(\frac{\partial f}{\partial \alpha} \right)$, $\frac{\partial}{\partial \alpha} \left(\frac{\partial f}{\partial \beta} \right)$, $\frac{\partial}{\partial \beta} \left(\frac{\partial f}{\partial \alpha} \right)$, and hence locate the maximum $\left(\frac{\partial f}{\partial \alpha}=0, \frac{\partial f}{\partial \beta}=0 \right)$.

We find in this way that the maximum occurs at $\alpha=0.784$, $\beta=0.510$, and that for these values

$$\frac{\sigma(y)}{\sigma(b_{22})} = 0.849613 \frac{(n^2-1)\sqrt{n}}{12}.$$

Now for the least-squares coefficient

$$\left[\frac{\sigma(y)}{\sigma(b'_{22})} \right]^2 = \frac{n(n^2-1)(n^2-4)}{180},$$

and so if n is large $\eta_1(b_{22})=90.2$ per cent.

Also

$$\frac{\sigma(y)}{\sigma(b_{20})} \doteq \frac{n}{\sqrt{n}} \frac{(1-\alpha^3-\beta^3)-(1-\alpha-\beta)}{[(1-\alpha^3-\beta^3)^2+(1-\alpha+\beta)-2(1-\alpha-\beta)(1-\alpha^3-\beta^3)]^{\frac{1}{2}}},$$

and as

$$\frac{\sigma(y)}{\sigma(b'_{20})} \doteq \frac{\sqrt{n}}{1.5},$$

$\eta_1(b_{20})$ is 94.5 per cent.

This scheme of grouping makes $\eta_1(b_{22})$ a maximum while retaining the original form for W_0 . If W_0 is varied, it is found that $\eta_1(b_{22})$ can be increased still further, but this increase is accompanied by a drop in $\eta_1(b_{20})$; or $\eta_1(b_{20})$ can be increased, but this increase is accompanied by a drop in $\eta_1(b_{22})$. Hence no advantage is to be obtained from varying W_0 . It would be possible to choose W_2 to maximize $\eta_1(b_{20})$ instead of $\eta_1(b_{22})$, but we have not adopted this scheme because the efficiencies of the coefficients of higher degree appear to be more sensitive to changes in the groups, and also drop more rapidly when the spacing becomes markedly non-uniform.

In constructing the second degree equation, then, we take the first $n_2 = 0.108n$ observations, omit the next $v_2 = 0.137n$ observations, subtract the central $n'_2 = 0.510n$ observations, omit the next v_2 observations, and add the final n_2 observations.

(c) *Polynomial of the Third Degree*

The equations determining b_{31} and b_{33} are

$$\left. \begin{aligned} b_{31}\Sigma W_1x + b_{33}\Sigma W_1x^3 &= \Sigma W_1y, \\ b_{31}\Sigma W_3x + b_{33}\Sigma W_3x^3 &= \Sigma W_3y. \end{aligned} \right\} \dots\dots\dots (13)$$

The equations for b_{32} and b_{30} will be the same as those for b_{22} and b_{20} ($p=2$), and the optimum functions W_2 and W_0 will be identical with those for the case $p=2$.

The standard errors of the coefficients will be given by the equations

$$\frac{\sigma(y)}{\sigma(b_{33})} = \frac{\Sigma W_1x\Sigma W_3x^3 - \Sigma W_1x^3\Sigma W_3x}{[\Sigma W_3^2(\Sigma W_1x)^2 + \Sigma W_1^2(\Sigma W_3x)^2 - 2\Sigma W_1W_3\Sigma W_1x\Sigma W_3x]^{\frac{1}{2}}}, \dots (14)$$

$$\frac{\sigma(y)}{\sigma(b_{31})} = \frac{\Sigma W_1x\Sigma W_3x^3 - \Sigma W_1x^3\Sigma W_3x}{[\Sigma W_1^2(\Sigma W_3x^3)^2 + \Sigma W_3^2(\Sigma W_1x^3)^2 - 2\Sigma W_1W_3\Sigma W_1x^3\Sigma W_3x^3]^{\frac{1}{2}}}, \dots\dots\dots (15)$$

$\Sigma W_1f(x)$, $\Sigma W_3f(x)$ can be expressed in the following forms :

$$\Sigma W_1f(x) = \left\{ \sum_{0, \frac{1}{2}}^{\frac{n-1}{2}} - \sum_{0, \frac{1}{2}}^{\frac{a-1}{2}} \right\} \{f(+x) - f(-x)\},$$

$$\Sigma W_3f(x) = \left\{ \sum_{0, \frac{1}{2}}^{\frac{n-1}{2}} - \sum_{0, \frac{1}{2}}^{\frac{b-1}{2}} - \sum_{0, \frac{1}{2}}^{\frac{c-1}{2}} + \sum_{0, \frac{1}{2}}^{\frac{d-1}{2}} \right\} \{f(+x) - f(-x)\}.$$

Hence

$$\Sigma W_1x^3 \doteq \frac{(n^2-1)^2}{32}(1-\alpha^4),$$

$$\Sigma W_3x^3 \doteq \frac{(n^2-1)^2}{32}(1-\beta^4-\gamma^4+\delta^4),$$

where $\alpha=a/n$, $\beta=b/n$, $\gamma=c/n$, and $\delta=d/n$, and so equation (14) becomes

$$\frac{\sigma(y)}{\sigma(b_{33})} \doteq \frac{(n^2-1)^2}{32\sqrt{n}} \frac{(1-\alpha^2)(1-\beta^4-\gamma^4+\delta^4) - (1-\alpha^4)(1-\beta^2-\gamma^2+\delta^2)}{\Delta_{33}^{\frac{1}{2}}}, \dots\dots\dots (16)$$

where

$$\Delta_{33} = (1-\alpha^2)^2(1-\beta+\gamma-\delta) + (1-\alpha)(1-\beta^2-\gamma^2+\delta^2)^2 \\ - 2(1-\alpha^2)(1-\beta^2-\gamma^2+\delta^2)(1-\beta-\gamma+\alpha).$$

The form of the last term in Δ_{33} depends on the relative values of γ , δ , and α ; the expression given here corresponds to the case $\gamma > \alpha > \delta$.

If we retain the same function W_1 as in the case $p=1$, $\alpha=0.333$, and the maximum of $\sigma(y)/\sigma(b_{33})$ occurs when $\beta=0.865$, $\gamma=0.728$, $\delta=0.162$, the value then being

$$0.564491 \frac{(n^2-1)^2}{32\sqrt{n}}.$$

But

$$\left[\frac{\sigma(y)}{\sigma(b'_{33})} \right]^2 = \frac{n(n^2-1)(n^2-4)(n^2-9)}{2800}$$

and hence $\eta(b_{33})$ is 87.1 per cent. For b_{31} we find

$$\frac{\sigma(y)}{\sigma(b_{31})} \doteq 0.441951 \frac{n^2-1}{4\sqrt{n}},$$

$$\left[\frac{\sigma(b'_{31})}{\sigma(y)} \right]^2 \doteq \frac{75}{n(n^2-1)},$$

and so $\eta(b_{31})$ is 91.6 per cent.

The efficiency of the coefficient b_{33} , although probably satisfactory, is a little low. Also experience shows that, when the observations are clustered towards one end of the range, the efficiencies of the higher degree coefficients drop rather sharply. Hence it is desirable to improve the efficiency by changing the function W_1 .

The maximum of the expression on the right-hand side of equation (16), with $\alpha, \beta, \gamma, \delta$, all variable, occurs when $\alpha=0.706, \beta=0.917, \gamma=0.831, \delta=0.125$. The value of $\sigma(y)/\sigma(b_{33})$ is then

$$0.585249 \frac{(n^2-1)^2}{32\sqrt{n}},$$

and hence $\eta(b_{33})$ is 93.7 per cent. Also

$$\frac{\sigma(y)}{\sigma(b_{31})} \doteq 0.443515 \frac{(n^2-1)}{4\sqrt{n}},$$

and so $\eta(b_{31})$ is 92.2 per cent. Thus altering W_1 produces a worthwhile improvement in the efficiency.

In fitting a polynomial of the third degree, then, we construct the first degree equation by taking the first $n_1=0.147n$ observations, omitting the central $v_1=0.706n$ observations, and subtracting the last n_1 observations. We construct the third degree equation by taking the first $n_3=0.0415n$ observations, omitting the next $v_3=0.043n$ observations, subtracting the next $n'_3=0.353n$ observations, omitting the central $v'_3=0.125n$ observations, adding the next n'_3 observations, omitting the next v_3 observations, and subtracting the last n_3 observations.

Occasionally it may be desirable to retain the first degree equation for $p=1$, for example, to compare the second and third degree polynomials to see which gives the better fit. The groups for the third degree equation can then be found from the formulae:

$$n_3=0.0675n,$$

$$v_3=0.0685n,$$

$$n'_3=0.283n,$$

$$v'_3=0.162n.$$

(d) *Polynomial of the Fourth Degree*

The equations determining b_{40} , b_{42} , b_{44} are

$$\left. \begin{aligned} b_{40}\Sigma W_0 + b_{42}\Sigma W_0x^2 + b_{44}\Sigma W_0x^4 &= \Sigma W_0y, \\ b_{40}\Sigma W_2 + b_{42}\Sigma W_2x^2 + b_{44}\Sigma W_2x^4 &= \Sigma W_2y, \\ b_{40}\Sigma W_4 + b_{42}\Sigma W_4x^2 + b_{44}\Sigma W_4x^4 &= \Sigma W_4y. \end{aligned} \right\} \dots (17)$$

The equations for b_{41} , b_{43} are identical with those for b_{31} , b_{33} in the case $p=3$, and the functions W_1 , W_3 will be those found in Section III (c).

The solutions of equations (17) may be written as follows :

$$\begin{aligned} b_{44}(\Sigma W_4x^4 + \lambda_{42}\Sigma W_2x^4 + \lambda_{40}\Sigma W_0x^4) &= \Sigma(W_4 + \lambda_{42}W_2 + \lambda_{40}W_0)y, \\ b_{42}(\Sigma W_2x^2 + \lambda_{24}\Sigma W_4x^2 + \lambda_{20}\Sigma W_0x^2) &= \Sigma(W_2 + \lambda_{24}W_4 + \lambda_{20}W_0)y, \\ b_{40}(\Sigma W_0 + \lambda_{02}\Sigma W_2 + \lambda_{04}\Sigma W_4) &= \Sigma(W_0 + \lambda_{02}W_2 + \lambda_{04}W_4)y, \end{aligned}$$

where

$$\begin{aligned} \lambda_{42} &= -(\Sigma W_4x^2\Sigma W_0 - \Sigma W_0x^2\Sigma W_4)/(\Sigma W_2x^2\Sigma W_0 - \Sigma W_0x^2\Sigma W_2), \\ \lambda_{40} &= -(\Sigma W_4 + \lambda_{42}\Sigma W_2)/\Sigma W_0, \\ \lambda_{24} &= -(\Sigma W_2x^4\Sigma W_0 - \Sigma W_0x^4\Sigma W_2)/(\Sigma W_4x^4\Sigma W_0 - \Sigma W_0x^4\Sigma W_4), \\ \lambda_{20} &= -(\Sigma W_2 + \lambda_{24}\Sigma W_4)/\Sigma W_0, \\ \lambda_{02} &= -(\Sigma W_0x^2\Sigma W_4x^4 - \Sigma W_4x^2\Sigma W_0x^4)/(\Sigma W_2x^2\Sigma W_4x^4 - \Sigma W_4x^2\Sigma W_2x^4), \\ \lambda_{04} &= -(\Sigma W_0x^4 + \lambda_{02}\Sigma W_2x^4)/\Sigma W_4x^4. \end{aligned}$$

Thus

$$\frac{\sigma(y)}{\sigma(b_{44})} = \frac{\Sigma W_4x^4 + \lambda_{42}\Sigma W_2x^4 + \lambda_{40}\Sigma W_0x^4}{\Delta_{44}^{\frac{1}{2}}},$$

where

$$\Delta_{44} = \Sigma W_4^2 + \lambda_{42}^2 \Sigma W_2^2 - \lambda_{40}^2 \Sigma W_0^2 + 2\lambda_{42}\lambda_{40}\Sigma W_0W_2 + 2\lambda_{42}\Sigma W_4W_2 + 2\lambda_{40}\Sigma W_4W_0.$$

Now we can write

$$\begin{aligned} \Sigma W_2f(x) &= \left\{ \begin{array}{ccc} \frac{n-1}{2} & \frac{a-1}{2} & \frac{b-1}{2} \\ \Sigma & - \Sigma & - \Sigma \end{array} \right\} f(x), \\ \Sigma W_4f(x) &= \left\{ \begin{array}{ccccc} \frac{n}{2} & \frac{c-1}{2} & \frac{d-1}{2} & \frac{e-1}{2} & \frac{f-1}{2} \\ \Sigma & - \Sigma & - \Sigma & + \Sigma & + \Sigma \end{array} \right\} f(x), \end{aligned}$$

where the parameters a, b, c, d, e, f locate the groups. If we set $a=\alpha n$, $b=\beta n$, $c=\gamma n$, $d=\delta n$, $e=\varepsilon n$, $f=\varphi n$, then

$$\Sigma W_2x^4 \doteq \frac{n(n^2-1)(3n^2-7)}{240}(1-\alpha^5-\beta^5),$$

$$\Sigma W_4x^4 \doteq \frac{n(n^2-1)(3n^2-7)}{240}(1-\gamma^5-\delta^5+\varepsilon^5+\varphi^5).$$

Suppose we retain the functions W_0 , W_2 derived for $p=2$. That is, we set $\alpha=0.784$, $\beta=0.510$. We find firstly that as ε is varied the function $\sigma(y)/\sigma(b_{44})$

has a maximum at $\varepsilon = \beta$. This comes about because the expression $\Sigma W_2 W_4$ contains an extra term $\beta - \varepsilon$ if $\beta > \varepsilon$. We accordingly set $\varepsilon = 0.510$, and determine the location of the maximum of $\sigma(y) \cdot \sigma(b_{44})$ as γ , δ , and φ are varied.

The maximum of this function occurs when $\gamma = 0.943$, $\delta = 0.876$, $\varepsilon = 0.510$, and $\varphi = 0.307$, the value then being

$$0.365014 \frac{n(n^2-1)(n^2-7.3)}{80}.$$

For the least-squares coefficient

$$\left[\frac{\sigma(b'_{44})}{\sigma(y)} \right]^2 = \frac{44100}{n(n^2-1)(n^2-4)(n^2-9)(n^2-16)},$$

and so the efficiency $\eta(b_{44})$ is 91.8 per cent.

For the coefficient b_{42}

$$\frac{\sigma(y)}{\sigma(b_{42})} \doteq 0.244578 \frac{\sqrt{n(n^2-1)}}{12},$$

$$\left[\frac{\sigma(b'_{42})}{\sigma(y)} \right]^2 \doteq \frac{2205}{n(n^2-1)(n^2-4)},$$

and so the efficiency $\eta(b_{42})$ is 91.6 per cent.

For the coefficient b_{40}

$$\frac{\sigma(y)}{\sigma(b_{40})} \doteq 0.515230 \sqrt{n},$$

$$\left[\frac{\sigma(b'_{40})}{\sigma(y)} \right]^2 \doteq \frac{3.515625}{n},$$

and so the efficiency $\eta(b_{40})$ is 93.3 per cent.

The efficiencies of these coefficients are all greater than 90 per cent., and so this method of grouping, in which the functions W_0 , W_2 derived for $p = 2$ are retained, will be quite satisfactory.

To form the fourth-degree equation, then, we take the first $n_4 = 0.0285n$ observations, omit the next $v_4 = 0.0335n$ observations, subtract the next $n'_4 = 0.183n$ observations, omit the next $v'_4 = 0.1015n$ observations, add the central $n''_4 = 0.307n$ observations, omit the next v'_4 observations, subtract the next n'_4 observations, omit the next v_4 observations, and add the final n_4 observations.

IV. TABLE OF GROUPS

The recommended methods of grouping are shown in Table 1. The observations are assumed to be listed in order of decreasing x . The numbers n_i give the number of observations in each group, while the numbers v_i give the number of observations omitted between groups. Table 2 is a summary of the efficiencies for large n . For smaller values of n the efficiencies tend to be somewhat higher.

The numbers n_j , v_j in Table 1 were calculated from the values obtained in Section III for the coefficients α , β , etc. However, the numbers n_j , v_j must be integers, and hence the numbers calculated were rounded off to the nearest integer to give the values in Table 1.

TABLE 1
GROUPING SCHEME

Zero degree equation: n

1st degree equation: $n_1(v_1) - n_1$

2nd degree equation: $n_2(v_2) - n'_2(v_2) + n_2$

3rd degree equation: $n_3(v_3) - n'_3(v'_3) + n'_3(v_3) - n_3$

4th degree equation: $n_4(v_4) - n'_4(v'_4) + n''_4(v'_4) - n'_4(v_4) + n_4$

n	$p < 2$		$p > 2$													
	n_1	v_1	n_1	v_1	n_2	v_2	n'_2	n_3	v_3	n'_3	v'_3	n_4	v_4	n'_4	v'_4	n''_4
10	3	4	2	6	1	1	6	1	1	2	2	1	0	2	1	2
11	4	3	2	7	1	2	5	1	1	3	1	1	0	2	1	3
12	4	4	2	8	1	2	6	1	1	3	2	1	0	2	1	4
13	4	5	2	9	1	2	7	1	1	4	1	1	0	2	2	3
14	5	4	2	10	1	2	8	1	1	4	2	1	0	3	2	2
15	5	5	2	11	2	2	7	1	1	5	1	1	0	3	1	5
16	5	6	3	10	2	2	8	1	1	5	2	1	0	3	2	4
17	6	5	3	11	2	2	9	1	1	5	3	1	0	3	2	5
18	6	6	3	12	2	2	10	1	1	6	2	1	0	3	2	6
19	6	7	3	13	2	3	9	1	1	6	3	1	0	4	2	5
20	7	6	3	14	2	3	10	1	1	7	2	1	0	4	2	6
21	7	7	3	15	2	3	11	1	1	7	3	1	0	4	2	7
22	7	8	3	16	2	3	12	1	1	8	2	1	1	3	3	6
23	8	7	3	17	2	4	11	1	1	8	3	1	1	4	2	7
24	8	8	4	16	3	3	12	1	1	8	4	1	1	4	2	8
25	8	9	4	17	3	3	13	1	1	9	3	1	1	4	3	7
26	9	8	4	18	3	3	14	1	1	9	4	1	1	4	3	8
27	9	9	4	19	3	4	13	1	1	10	3	1	1	5	2	9
28	9	10	4	20	3	4	14	1	1	10	4	1	1	5	3	8
29	10	9	4	21	3	4	15	1	1	11	3	1	1	5	3	9
30	10	10	4	22	3	4	16	1	2	10	4	1	1	5	3	10
31	10	11	5	21	3	5	15	1	2	11	3	1	1	6	3	9
32	11	10	5	22	3	5	16	1	2	11	4	1	1	6	3	10
33	11	11	5	23	4	4	17	1	2	11	5	1	1	6	3	11
34	11	12	5	24	4	4	18	1	2	12	4	1	1	6	4	10
35	12	11	5	25	4	5	17	1	2	12	5	1	1	7	3	11
36	12	12	5	26	4	5	18	1	2	13	4	1	1	7	3	12
37	12	13	5	27	4	5	19	2	1	13	5	1	1	7	4	11
38	13	12	6	26	4	5	20	2	1	14	4	1	1	7	4	12
39	13	13	6	27	4	6	19	2	1	14	5	1	1	8	4	11
40	13	14	6	28	4	6	20	2	1	14	6	1	1	8	4	12

TABLE 1 (Continued)

n	p n_1	2 v_1	p n_1	2 v_1	n_2	v_2	n'_2	n_3	v_3	n'_3	v'_3	n_4	v_4	n'_4	v'_4	n''_4
41	14	13	6	29	4	6	21	2	1	15	5	1	2	7	4	13
42	14	14	6	30	5	5	22	2	2	14	6	1	2	7	5	12
43	14	15	6	31	5	6	21	2	2	15	5	1	2	8	4	13
44	15	14	6	32	5	6	22	2	2	15	6	1	2	8	4	14
45	15	15	7	31	5	6	23	2	2	16	5	1	2	8	5	13
46	15	16	7	32	5	6	24	2	2	16	6	1	2	8	5	14
47	16	15	7	33	5	7	23	2	2	17	5	1	2	9	4	15
48	16	16	7	34	5	7	24	2	2	17	6	1	2	9	5	14
49	16	17	7	35	5	7	25	2	2	17	7	1	2	9	5	15
50	17	16	7	36	5	7	26	2	2	18	6	1	2	9	5	16
51	17	17	7	37	6	6	27	2	2	18	7	1	2	9	6	15
52	17	18	8	36	6	7	26	2	2	19	6	1	2	10	5	16
53	18	17	8	37	6	7	27	2	2	19	7	2	1	10	5	17
54	18	18	8	38	6	7	28	2	3	19	6	2	1	10	6	16
55	18	19	8	39	6	7	29	2	3	19	7	2	1	10	6	17
56	19	18	8	40	6	8	28	2	3	19	8	2	1	11	5	18
57	19	19	8	41	6	8	29	2	3	20	7	2	2	10	6	17
58	19	20	9	40	6	8	30	2	3	20	8	2	2	10	6	18
59	20	19	9	41	6	8	31	2	3	21	7	2	2	10	6	19
60	20	20	9	42	6	9	30	2	3	21	8	2	2	11	6	18
61	20	21	9	43	7	8	31	3	2	22	7	2	2	11	6	19
62	21	20	9	44	7	8	32	3	2	22	8	2	2	11	6	20
63	21	21	9	45	7	8	33	3	2	23	7	2	2	11	7	19
64	21	22	9	46	7	9	32	3	2	23	8	2	2	12	6	20
65	22	21	10	45	7	9	33	3	2	23	9	2	2	12	7	19
66	22	22	10	46	7	9	34	3	3	23	8	2	2	12	7	20
67	22	23	10	47	7	9	35	3	3	23	9	2	2	12	7	21
68	23	22	10	48	7	10	34	3	3	24	8	2	2	13	7	20
69	23	23	10	49	7	10	35	3	3	24	9	2	2	13	7	21
70	23	24	10	50	8	9	36	3	3	25	8	2	2	13	7	22
71	24	23	10	51	8	9	37	3	3	25	9	2	2	13	8	21
72	24	24	11	50	8	10	36	3	3	25	10	2	2	14	7	22
73	24	25	11	51	8	10	37	3	3	26	9	2	3	13	7	23
74	25	24	11	52	8	10	38	3	3	26	10	2	3	13	8	22
75	25	25	11	53	8	10	39	3	3	27	9	2	3	13	8	23
76	25	26	11	54	8	11	38	3	3	27	10	2	3	14	7	24
77	26	25	11	55	8	11	39	3	4	27	9	2	3	14	8	23
78	26	26	11	56	8	11	40	3	4	27	10	2	3	14	8	24
79	26	27	12	55	9	10	41	3	4	28	9	2	3	14	8	25
80	27	26	12	56	9	11	40	3	4	28	10	2	3	15	8	24
81	27	27	12	57	9	11	41	3	4	28	11	2	3	15	8	25
82	27	28	12	58	9	11	42	3	4	29	10	2	3	15	8	26
83	28	27	12	59	9	11	43	3	4	29	11	2	3	15	9	25
84	28	28	12	60	9	12	42	3	4	30	10	2	3	16	8	26
85	28	29	12	61	9	12	43	4	3	30	11	2	3	16	8	27

TABLE 1 (Continued)

n	$p < 2$		$p > 2$													
	n_1	v_1	n_1	v_1	n_2	v_2	n'_2	n_3	v_3	n'_3	v'_3	n_4	v_4	n'_4	v'_4	n''_4
86	29	28	13	60	9	12	44	4	3	31	10	2	3	16	9	26
87	29	29	13	61	9	12	45	4	3	31	11	2	3	16	9	27
88	29	30	13	62	10	12	44	4	3	31	12	3	2	17	8	28
89	30	29	13	63	10	12	45	4	4	31	11	3	3	16	9	27
90	30	30	13	64	10	12	46	4	4	31	12	3	3	16	9	28
91	30	31	13	65	10	12	47	4	4	32	11	3	3	16	10	27
92	31	30	14	64	10	13	46	4	4	32	12	3	3	17	9	28
93	31	31	14	65	10	13	47	4	4	33	11	3	3	17	9	29
94	31	32	14	66	10	13	48	4	4	33	12	3	3	17	10	28
95	32	31	14	67	10	13	49	4	4	34	11	3	3	17	10	29
96	32	32	14	68	10	14	48	4	4	34	12	3	3	18	9	30
97	32	33	14	69	10	14	49	4	4	34	13	3	3	18	10	29
98	33	32	14	70	11	13	50	4	4	35	12	3	3	18	10	30
99	33	33	15	69	11	13	51	4	4	35	13	3	3	18	10	31
100	33	34	15	70	11	14	50	4	4	36	12	3	3	19	10	30

It is found that slight variations in the numbers in each group produce only a very small change in the efficiency, and so this procedure will be quite satisfactory provided n is not too small. Tables 3A, 3B, and 3C show the calculated efficiencies for the recommended methods of grouping at selected values of n . In these tables the groups are identified by the range of x —for example, (37-47) indicates that the observations lying between $x=37$ and $x=47$ (inclusive) are grouped together.

TABLE 2

EFFICIENCIES OF POLYNOMIAL COEFFICIENTS

b_{00}, b_{10}	100%				
b_{11}, b_{21}	88.9%				
b_{22}, b_{32}	90.2%	b_{20}, b_{30}	94.5%		
b_{33}, b_{43}	93.7%	b_{31}, b_{41}	92.2%		
b_{44}	91.8%	b_{42}	91.6%	b_{40}	93.3%

If n is small, the calculated groups may not yield satisfactory values for the efficiencies. The efficiencies at all values of n below 25 have been worked out individually, and the calculated groups altered where necessary to improve the efficiencies.

In deriving Table 1, the origin was assumed to be at the centre of the range of x , since for the equally-spaced case the coefficients of even degree are then independent of the coefficients of odd degree. In applying the method of grouping to an example, the origin may be taken at any convenient point.

TABLE 3A
EFFICIENCIES AT SELECTED VALUES OF n , $p=2$

W	ΣW^2	ΣW	$\Sigma W'x^2$	$\sigma(y)$ $\sigma(b_{22})$	$\sigma(y)$ $\sigma(b_{20})$
$n=100$				Least-squares	
W_0 ($\frac{1}{2}$ -49 $\frac{1}{2}$)	100	100	83325	7451.7	6.6661
W_2 ($39\frac{1}{2}$ -49 $\frac{1}{2}$)-($\frac{1}{2}$ -24 $\frac{1}{2}$)	72	-28	33373	Grouping	
				7079.2	6.4746
				(90.3%)	(94.3%)
$n=95$				Least-squares	
W_0 (0-47)	95	95	71440	6554.7	6.4972
W_2 (38-47)-(0-24)	69	-29	26490	Grouping	
				6227.6	6.3110
				(90.3%)	(94.4%)
$n=80$				Least-squares	
W_0 ($\frac{1}{2}$ -39 $\frac{1}{2}$)	80	80	42660	4265.0	5.9620
W_2 ($31\frac{1}{2}$ -39 $\frac{1}{2}$)-($\frac{1}{2}$ -19 $\frac{1}{2}$)	58	-22	17474.5	Grouping	
				4052.1	5.7911
				(90.3%)	(94.3%)
$n=75$				Least-squares	
W_0 (0-37)	75	75	35150	3629.3	5.7726
W_2 (30-37)-(0-19)	55	-23	13100	Grouping	
				3448.6	5.6076
				(90.3%)	(94.4%)
$n=60$				Least-squares	
W_0 ($\frac{1}{2}$ -29 $\frac{1}{2}$)	60	60	17995	2077.0	5.1627
W_2 ($24\frac{1}{2}$ -29 $\frac{1}{2}$)-($\frac{1}{2}$ -14 $\frac{1}{2}$)	42	-18	6535.5	Grouping	
				1972.6	5.0136
				(90.2%)	(94.3%)
$n=48$				Least-squares	
W_0 ($\frac{1}{2}$ -23 $\frac{1}{2}$)	48	48	9212	1188.5	4.6171
W_2 ($19\frac{1}{2}$ -23 $\frac{1}{2}$)-($\frac{1}{2}$ -11 $\frac{1}{2}$)	34	-14	3492.5	Grouping	
				1129.8	4.4860
				(90.4%)	(94.4%)
$n=36$				Least-squares	
W_0 ($\frac{1}{2}$ -17 $\frac{1}{2}$)	36	36	3885	578.47	3.9974
W_2 ($14\frac{1}{2}$ -17 $\frac{1}{2}$)-($\frac{1}{2}$ -8 $\frac{1}{2}$)	26	-10	1573.5	Grouping	
				550.47	3.8862
				(90.6%)	(94.5%)
$n=20$				Least-squares	
W_0 ($\frac{1}{2}$ -9 $\frac{1}{2}$)	20	20	665	132.50	2.9752
W_2 ($8\frac{1}{2}$ -9 $\frac{1}{2}$)-($\frac{1}{2}$ -4 $\frac{1}{2}$)	14	-6	242.5	Grouping	
				126.54	2.8984
				(91.2%)	(94.9%)

TABLE 3B
EFFICIENCIES AT SELECTED VALUES OF n , $p=3$

W	ΣW^2	$\Sigma W_1 W_2$	ΣWx	ΣWx^2	$\frac{\sigma(y)}{\sigma(b_{33})}$ ($\times 10^{-3}$)	$\frac{\sigma(y)}{\sigma(b_{31})}$
$n=100$					Least-squares	
W_1 (35 $\frac{1}{2}$ -49 $\frac{1}{2}$)	36		1275	2374368.75	188.85	115.42
W_2 (46 $\frac{1}{2}$ -49 $\frac{1}{2}$)-(6 $\frac{1}{2}$ -41 $\frac{1}{2}$)	80	-6	-1344	-668592	Grouping	
					182.80	110.82
					(93.7%)	(92.2%)
$n=95$					Least-squares	
W_1 (34-47)	28		1134	1915326	157.80	106.87
W_2 (44-47)-(6-39)	76	4	-1166	-461414	Grouping	
					152.81	102.62
					(93.8%)	(92.2%)
$n=80$					Least-squares	
W_1 (28 $\frac{1}{2}$ -39 $\frac{1}{2}$)	24		816	972468	86.449	82.572
W_2 (37 $\frac{1}{2}$ -39 $\frac{1}{2}$)-(5 $\frac{1}{2}$ -32 $\frac{1}{2}$)	62	-4	-833	-249520.25	Grouping	
					83.700	79.504
					(93.7%)	(92.7%)
$n=75$					Least-squares	
W_1 (27-37)	22		704	742016	68.960	74.946
W_2 (35-37)-(5-31)	60	4	-756	-211464	Grouping	
					66.804	71.994
					(93.8%)	(92.3%)
$n=60$					Least-squares	
W_1 (21 $\frac{1}{2}$ -29 $\frac{1}{2}$)	18		459	307644.75	31.558	53.606
W_2 (28 $\frac{1}{2}$ -29 $\frac{1}{2}$)-(4 $\frac{1}{2}$ -24 $\frac{1}{2}$)	46	-4	-493	-97389.25	Grouping	
					30.555	51.676
					(93.7%)	(92.9%)
$n=48$					Least-squares	
W_1 (17 $\frac{1}{2}$ -23 $\frac{1}{2}$)	14		287	124055.75	14.436	38.333
W_2 (22 $\frac{1}{2}$ -23 $\frac{1}{2}$)-(3 $\frac{1}{2}$ -19 $\frac{1}{2}$)	38	-2	-299	-31124.75	Grouping	
					14.013	36.883
					(94.2%)	(92.6%)
$n=36$					Least-squares	
W_1 (13 $\frac{1}{2}$ -17 $\frac{1}{2}$)	10		155	38168.75	5.2617	24.864
W_2 (17 $\frac{1}{2}$)-(2 $\frac{1}{2}$ -14 $\frac{1}{2}$)	28	-2	-186	-14530.5	Grouping	
					5.0999	24.032
					(93.9%)	(93.4%)
$n=20$					Least-squares	
W_1 (7 $\frac{1}{2}$ -9 $\frac{1}{2}$)	6		51	3786.75	0.66429	10.224
W_2 (9 $\frac{1}{2}$)-(1 $\frac{1}{2}$ -7 $\frac{1}{2}$)	16	0	-44	-317	Grouping	
					0.65209	9.920
					(96.4%)	(94.1%)

TABLE 3C
EFFICIENCIES AT SELECTED VALUES OF $n, p=4$

W	ΣW^2	ΣW	$\Sigma W_2 W_4$	$\Sigma W x^2$	$\Sigma W x^4$	$\frac{\sigma(y)}{\sigma(b_{44})}$ ($\times 10^{-2}$)	$\frac{\sigma(y)}{\sigma(b_{42})}$	$\frac{\sigma(y)}{\sigma(b_{40})}$
$n=100$								
W_0 (1-49½)	100	100		83325	124958336	Least-squares		
W_2 (39½-49½)	72	-28	-34	33373	84987384	47548	2127.4	5.3321
-(½-24½)						Grouping		
W_4 (47½-49½)	74	-2		-30004.5	-28479490	45633	2039.5	5.1557
-(25½-43½)						(92.1%)	(91.9%)	(93.5%)
(½-14½)								
$n=95$								
W_0 (0-47)	95	95		71440	96686896	Least-squares		
W_2 (38-47)						37741	1871.2	5.1969
-(0-24)	69	-29	-31	26490	63515346	Grouping		
W_4 (45-47)						36190	1790.3	5.0145
-(25-41)	69	+1		-23112	-18517248	(91.9%)	(91.5%)	(93.1%)
(0-14)								
$n=80$								
W_0 (½-39½)	80	80		42660	40938669	Least-squares		
W_2 (31½-39½)						17405	1217.1	4.7685
-(½-19½)	58	-22	-28	17474.5	28219603	Grouping		
W_4 (38½-39½)						16707	1168.7	4.6198
-(20½-34½)	58	-2		-16012.5	-10355278	(92.1%)	(92.2%)	(93.9%)
(½-11½)								
$n=75$								
W_0 (0-37)	75	75		35150	29645510	Least-squares		
W_2 (30-37)						13013	1035.6	4.6169
-(0-19)	55	-23	-25	13100	19592180	Grouping		
W_4 (36-37)						12479	991.88	4.4613
-(20-32)	53	+1		-11598	-6179358	(92.0%)	(91.7%)	(93.4%)
+(0-11)								
$n=60$								
W_0 (½-29½)	60	60		17995	9711002	Least-squares		
W_2 (24½-29½)						4760.4	592.20	4.1285
-(½-14½)	42	-18	-18	6535.5	6227933	Grouping		
W_4 (28½-29½)						4578.9	568.69	3.9925
-(15½-25½)	44	0		-5616	-1586520	(92.5%)	(92.2%)	(93.5%)
+(½-8½)								

TABLE 3C (Continued)

W	ΣW^2	ΣW	$\Sigma W_2 W_4$	$\Sigma W x^2$	$\Sigma W x^4$	$\frac{\sigma(y)}{\sigma(b_{44})}$ ($\times 10^{-2}$)	$\frac{\sigma(y)}{\sigma(b_{42})}$	$\frac{\sigma(y)}{\sigma(b_{40})}$
$n=48$								
W_0 ($\frac{1}{2}-23\frac{1}{2}$)	48	48		9212	3180443	Least-squares		
W_2 ($19\frac{1}{2}-23\frac{1}{2}$)						1739.9	338.46	3.6913
—($\frac{1}{2}-11\frac{1}{2}$)	34	-14	-16	3492.5	2093331.1	Grouping		
W_4 ($23\frac{1}{2}$)						1675.0	325.94	3.5812
—($12\frac{1}{2}-20\frac{1}{2}$)	34	-2		-3688.5	-915028.1	(92.7%)	(92.7%)	(94.1%)
—($\frac{1}{2}-6\frac{1}{2}$)								
$n=36$								
W_0 ($\frac{1}{2}-17\frac{1}{2}$)	36	36		3885	753884.25	Least-squares		
W_2 ($14\frac{1}{2}-17\frac{1}{2}$)						474.34	164.32	3.1942
—($\frac{1}{2}-8\frac{1}{2}$)	26	-10	-14	1573.5	516291.38	Grouping		
W_4 ($17\frac{1}{2}$)						458.54	158.34	3.0921
—($9\frac{1}{2}-15\frac{1}{2}$)	28	0		-1488	-204072	(93.4%)	(92.9%)	(93.7%)
—($\frac{1}{2}-5\frac{1}{2}$)								
$n=20$								
W_0 ($\frac{1}{2}-9\frac{1}{2}$)	20	20		665	39667.25	Least-squares		
W_2 ($8\frac{1}{2}-9\frac{1}{2}$)						32.801	37.136	2.3709
—($\frac{1}{2}-4\frac{1}{2}$)	14	-6	-6	242.5	25521.625	Grouping		
W_4 ($9\frac{1}{2}$)						31.771	35.693	2.2846
—($5\frac{1}{2}-8\frac{1}{2}$)	16	0		-204	-5790	(93.8%)	(92.4%)	(92.9%)
—($\frac{1}{2}-2\frac{1}{2}$)								

When the observations are not equally spaced, the coefficients obtained by the method of grouping will still be unbiased, but the efficiencies may be somewhat lower. An irregular variation throughout the range in the interval between successive observations will not affect the efficiencies greatly, and the method can be used with confidence in such cases. However, in cases where the observations are clustered at one end of the range, the efficiencies may drop below the values quoted in Table 2. It is not possible to give any definite limiting values for the efficiencies, but in specific examples which have been worked out the efficiencies of the coefficients of higher powers of x have been found to drop to about 80 per cent. for the quadratic and cubic, and to about 70 per cent. for the quartic.

V. ILLUSTRATIVE EXAMPLE

We consider the fitting of a cubic to the 16 observations listed in Model Form. 1, the observations being numbered in order of decreasing x .

MODEL FORM 1

1st : $\Sigma(3) + \Sigma(13) - \Sigma(16)$ 2nd : $\Sigma(2) + \Sigma(4) - \Sigma(12) - \Sigma(14) - \Sigma(16)$ 3rd : $\Sigma(1) + \Sigma(2) - \Sigma(7) - \Sigma(9) + \Sigma(14) + \Sigma(15) - \Sigma(16)$

No.	y		x		x^2		x^3	
1	8.10	8.10	+7.9	7.9	62.41	62.41	+493.039	493.039
2	6.85	14.95	+7.2	15.1	51.84	114.25	+373.248	866.287
3	5.08	20.03	+6.25	21.35	39.0625	153.3125	+244.140625	1110.427625
4	6.40	26.43	+6.1	27.45	37.21	190.5225	+226.981	1337.408625
5	4.50		+4.8		23.04		+110.592	
6	3.62		+3.5		12.25		+42.875	
7	2.85	37.40	+2.8	38.55	7.84	233.6525	+21.952	1512.827625
8	2.30		+0.8		0.64		+0.512	
9	1.45	41.15	-1.25	38.10	1.5625	235.8550	-1.953125	1511.386500
10	1.16		-2.2		4.84		-10.648	
11	1.17		-3.65		13.3225		-48.627125	
12	0.80	44.28	-4.4	27.85	19.36	273.3775	-85.184	1366.927375
13	0.76	45.04	-5.5	22.35	30.25	303.6275	-166.375	1200.552375
14	1.09	46.13	-6.4	15.95	40.96	344.5875	-262.144	938.408375
15	0.53	46.66	-7.3	8.65	53.29	397.8775	-389.017	549.391375
16	0.40	47.06	-9.0	-0.35	81.00	478.8775	-729.000	-179.608625

A convenient method of determining the "normal" equations is to add each column and record the progressive totals. This method reduces the chance of error, and will bring about a considerable saving in time if n is large. Thus if $\Sigma(a)$ denotes the sum of the first a terms, the "normal" equations can be obtained from the following expressions :

Zero degree : $\Sigma(16)$,1st degree : $\Sigma(3) + \Sigma(13) - \Sigma(16)$,2nd degree : $\Sigma(2) + \Sigma(4) - \Sigma(12) - \Sigma(14) + \Sigma(16)$,3rd degree : $\Sigma(1) + \Sigma(2) - \Sigma(7) - \Sigma(9) + \Sigma(14) + \Sigma(15) - \Sigma(16)$.

The numbers in brackets are n_i , n_j , ..., $n_i - n_j - n'_i$, etc. Note that pairs of positive signs and pairs of negative signs alternate. Lines are drawn across the Model Form at points where the progressive totals are required, and the values entered as the column is added.

In the present example the equations are

$$\begin{aligned} 16b_0 - 0.35b_1 + 478.8775b_2 - 179.608625b_3 &= 47.06, \\ +44.05b_1 - 21.9375b_2 + 2490.588625b_3 &= 18.01, \\ -4b_0 - 1.60b_1 + 165.6850b_2 - 281.248750b_3 &= -1.97, \\ -28.70b_1 - 29.2600b_2 + 2.520250b_3 &= -9.77. \end{aligned}$$

The values obtained by solving these equations are shown in Table 4, together with the least-squares values and their standard errors. The efficiencies of the coefficients obtained by grouping have been worked out (by expressing b_j as a linear function of the W_j) and are also included in the table. They are quite close to the values for the equally-spaced case given in Table 2.

TABLE 4

Least-Squares Solution			Grouping Method Solution	
	b_{aj}	S.E.	b_{aj}	Efficiency
b_{30}	1.80 ± 0.15		1.82	93%
$10^1 b_{31}$	3.14 ± 0.42		3.01	92%
$10^2 b_{32}$	3.91 ± 0.41		3.87	90%
$10^3 b_{33}$	2.40 ± 0.82		2.24	90%

VI. ESTIMATION OF ERRORS

We first derive an expression for the standard error of an observation in terms of the residuals.

Quantities occurring in the least-squares solution will be identified by a prime superscript. From least-squares theory we have

$$b'_{pk} = \sum_{r=k}^p \beta'_{rk} a'_r,$$

where

$$a'_r = \frac{\sum T_r(x)y(x)}{\sum T_r^2},$$

and

$$T_r(x) = \sum_{q=0}^r \beta'_{rq} x^q,$$

$T_r(x)$ being the orthogonal polynomial of degree r in x .

We show first that, if f is any linear function of the polynomial coefficients b_{pk} ,

$$\sigma^2(f-f') = \sigma^2(f) - \sigma^2(f'), \quad \dots \dots \dots (18)$$

where $\sigma(f)$ is the standard error of the function f . For

$$\begin{aligned} \sigma^2(f-f') &= E(f-f')^2, \\ &= E(f^2) + E(f'^2) - 2E(ff'). \end{aligned}$$

Now ff' is made up of terms $b_{pj}b'_{pk}$, and

$$\begin{aligned} E[b_{pj}b'_{pk}] &= \sum_x \left[\frac{w_{pj}(x)}{\sum w_{pj}x^j} \left(\sum_{r=k}^p \beta'_{rk} \frac{T_r(x)}{\sum T_r^2} \right) \right] \sigma^2(y), \\ &= \left[\sum_r \frac{\beta'_{rj}\beta'_{rk}}{\sum T_r^2} \right] \sigma^2(y), \\ &= E[b'_{pj}b'_{pk}]. \end{aligned}$$

Thus

$$E(ff') = E(f'^2),$$

and hence equation (18) follows.

The polynomial can be written in the form

$$u_p(x) = \sum_{j=0}^p a_j T_j(x),$$

where a_j is a linear function of the coefficients b_{pj} . Hence

$$\begin{aligned} E(a_j - a'_j)^2 &= \sigma^2(a_j) - \sigma^2(a'_j), \\ &= \{(1/\tau_{jj}) - 1\} \sigma^2(y) \sum T_j^2, \end{aligned}$$

where τ_{jj} is the efficiency for the determination of the coefficient a_j . If we denote the residuals by the symbol $v_p(x)$,

$$\begin{aligned} E[\sum r_p^2 - \sum r'_p{}^2] &= E \left[\sum_x (y - \sum_{j=0}^p a_j T_j)^2 - \sum_x (y - \sum_{j=0}^p a'_j T_j)^2 \right], \\ &= \sum_j E(a_j - a'_j)^2 \sum T_j^2, \\ &= \sum_j \{(1/\tau_{jj}) - 1\} \sigma^2(y), \end{aligned}$$

and as

$$E[\sum v_p^2] = (n - p - 1) \sigma^2(y),$$

it follows that

$$E[\sum r_p^2] = [n - (p + 1) + \sum \{(1/\tau_{jj}) - 1\}] \sigma^2(y).$$

In the method of grouping the efficiencies of the coefficients b_{pj} are all close to 90 per cent. Hence we can to a good approximation set

$$(1/\tau_{jj}) - 1 = 1 - \tau_{jj} = 0.1,$$

and so

$$E[\sum v_p^2] = [n - 0.9(p + 1)] \sigma^2(y).$$

Thus the formula

$$s_p = \left[\frac{\sum r_p^2}{n - 0.9(p + 1)} \right]^{1/2},$$

will provide an estimate of the standard error of an observation.

To find an approximation to the errors of the coefficients, we can use the functions φ_{pj} tabulated for the equally-spaced case (Guest 1950). Thus if the range of x is from x_1 to x_n , the standard error of the coefficient b_{pj} for the polynomial with origin at $x = \xi$ can be estimated from the formula

$$s[b_{pj}(k, n)] \approx \frac{1.05 \varphi_{pj}(k, n)}{h^j n^{j+\frac{1}{2}}} s_p,$$

where

$$h = \left| \frac{x_1 - x_n}{n - 1} \right|, \quad k = \frac{2\xi - (x_1 + x_n)}{nh},$$

the factor 1.05 allowing for the efficiency of 90 per cent.

This approximate formula should give values of $s[b_{pj}]$ quite close to the exact values when the spacing is roughly uniform. If the observations are clustered towards one end of the range, however, the discrepancy between the calculated and exact values may be greater. This discrepancy has never been observed to exceed 25 per cent. in the various examples which have been worked out.

VII. REFERENCES

- GUEST, P. G. (1950).—*Aust. J. Sci. Res. A* **3**: 173, 364.
 GUEST, P. G. (1951).—*Ann. Math. Statist.* **22**: 537.
 JEFFREYS, H. (1948).—"Theory of Probability." 2nd Ed. (Oxford Univ. Press.)
 NAIR, K. R., and SHRIVASTAVA, M. P. (1942).—*Sankhya* **6**: 121.

THE SCATTERING OF FAST ELECTRONS AND POSITRONS BY HEAVY ELEMENTS

By E. M. GUNNERSEN*

[Manuscript received December 19, 1951]

Summary

The angular distribution for the single elastic scattering of 1.07 MeV. positrons and electrons by the Hartree field of mercury ($Z = 80$) is calculated. The results show no large deviations from Coulomb values except, possibly, a slight increase in the region of 90° for electrons.

I. INTRODUCTION

Any accurate measurements of the single elastic scattering of fast electrons by heavy elements cannot be satisfactorily compared with theory unless one takes account of the screening effects of the extra-nuclear electrons. However, the only such calculations which have been published are those of Bartlett and Welton (1940), Massey and Mohr (1941), and Mohr (1943). The angular distributions obtained by these authors show very marked divergence for scattering angles greater than 60° .

No corresponding calculation for positrons has so far appeared.

II. GENERAL THEORY

The differential cross section $I(\theta)$ is given by the well-known Mott scattering equations (Mott and Massey 1949)

$$I(\theta) = |f|^2 + |g|^2,$$

where

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} \{ (l+1)(e^{2i\delta_l} - 1) + l(e^{2i\delta_{-l-1}} - 1) \} P_l(\cos \theta),$$

$$g(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} \{ -e^{2i\delta_l} + e^{2i\delta_{-l-1}} \} P_l'(\cos \theta),$$

and where $k^2 = 4\pi^2(E^2 - m_0^2 c^4)/h^2 c^2$, E being the total energy.

The δ_l are the phase differences at infinity between

$$(\frac{1}{2}\pi kr)^{\frac{1}{2}} J_{l+\frac{1}{2}}(kr),$$

and that bounded solution $M_l(r)$ of the equation

$$\frac{d^2 M_l}{dr^2} + \left\{ k^2 - \frac{l(l+1)}{r^2} - U_l(r) \right\} M_l = 0,$$

which vanishes at the origin.

* Physics Department, University of Melbourne.

U_l is the modified Dirac potential

$$U_l(r) = \frac{4\pi E}{hc^2} eV - \frac{4\pi^2}{h^2} \cdot \frac{e^2 V^2}{c^2} - \frac{l+1}{r} \cdot \frac{\alpha'}{\alpha} + \frac{3}{4} \cdot \frac{\alpha'^2}{\alpha^2} - \frac{1}{2} \cdot \frac{\alpha''}{\alpha},$$

where

$$\alpha = \frac{2\pi}{h} \left[\frac{E}{c} - \frac{eV}{c} - m_0 c \right], \quad \alpha' = \frac{2\pi}{h} \left[\frac{E}{c} - \frac{eV}{c} - m_0 c \right], \quad \alpha'' = d\alpha/dr, \text{ etc.}$$

The value used for the scattering potential $V(r)$ was that given by Hartree and Hackett (1935) for mercury ($Z=80$), namely $V(r) = -Z_p(r)/r$. Thus for positrons, the calculation was effected by changing Z_p to $-Z_p$ throughout.

The tabulation of $U_l(r)$ for different l and r was simplified by transforming $U_l(r)$ into the form

$$U_l(x) = Y + l \cdot X, \quad x = r/a_0,$$

where Y and X are functions of x . Z_p , dZ_p/dx , $d^2Z_p/dx^2/a_0$ is, as usual, the radius of the first Bohr orbit of hydrogen). To evaluate the derivatives of Z_p , use was made of Bartlett and Welton's empirical equations:

$$2Z_p = 33.7e^{-17.375x} + 95.9e^{-3.222x} - 30.3e^{-1.253x} \text{ for } x < 0.02,$$

$$2Z_p = 160e^{-6.325x} \text{ for } x < 0.02.$$

The whole calculation was performed for an incident K.E. of 1068.2 keV., for which $ka_0 = 400$.

III. DETERMINATION OF PHASE SHIFTS

(a) Low Order Phases

If x_0 is a zero of the function $M_l(x)$, then the contribution, $\delta_l|_{x_0}^{x_0}$, to the phase shift δ_l between the origin and x_0 is given by the equation

$$[-\pi ka_0] \cos \delta_l - \frac{2}{\pi} J_{l+1/2}(ka_0 x_0) - (-1)^l \sin \delta_l - \frac{2}{\pi} J_{l-1/2}(ka_0 x_0) = 0.$$

In order to determine x_0 , it is necessary to tabulate the modified wave function $M_l(x)$ by numerically integrating the equation

$$\frac{d^2 M_l}{dx^2} + f_l(x) \cdot M_l = 0,$$

where

$$f_l(x) = (ka_0)^2 - \frac{l(l+1)}{x^2} - U_l(x).$$

These numerical integrations—of which there were 26—were carried out by using Fox and Goodwin's (1949) recursion formula, viz.

$$\left(1 - \frac{1}{12} h^2 f_n\right) M_n - \left(2 - \frac{5}{6} h^2 f_{n-1}\right) M_{n-1} - \left(1 - \frac{1}{12} h^2 f_{n-2}\right) M_{n-2} + \Delta_n,$$

where

$$\Delta_n = \left(-\frac{1}{240} \delta^2 + \frac{13}{15120} \delta^3 - \dots\right) M_{n-1},$$

δ being the x -interval, and δ , δ^2 , ... being the usual 1st, 2nd, ... differences in M .

To avoid the construction of difference tables, h was made so small that Δ_n could be neglected. The actual x -ranges used were

$$0(0.0002)0.01, 0.01(0.0004)x_0.$$

In this way, the contributions, $\delta_l |_{x_0}^{x_0}$, to the phases $\delta_{-4}, \delta_{-3}, \dots, \delta_6, \delta_7, \delta_9$ were determined to within an estimated error of ± 0.001 radians.

Beyond the point x_0 , the contributions, $\delta_l |_{x_0}^{\infty}$, to the phases were evaluated by means of Jeffrey's approximation (subsequently denoted by J.A.)

$$\begin{aligned} \delta_l \Big|_{x_0}^{\infty} &= \int_{x_0}^{\infty} \left\{ (ka_0)^2 - U_l(x) - \frac{(l+\frac{1}{2})^2}{x^2} \right\}^{\frac{1}{2}} dx - \int_{x_0}^{\infty} \left\{ (ka_0)^2 - \frac{(l+\frac{1}{2})^2}{x^2} \right\}^{\frac{1}{2}} dx, \\ &= \int_{x_0}^{\infty} F_1^{\frac{1}{2}} dx - \int_{x_0}^{\infty} F_2^{\frac{1}{2}} dx, \text{ say.} \end{aligned}$$

The validity of this method was satisfactorily established by comparing values of the contributions, $\delta_l |_{x_1}^{x_2}$, between the points x_1 and x_2 as calculated by J.A., namely

$$\delta_l \Big|_{x_1}^{x_2} = \int_{x_1}^{x_2} F_1^{\frac{1}{2}} dx - \int_{x_1}^{x_2} F_2^{\frac{1}{2}} dx,$$

with the corresponding differences as calculated by numerical integration. The results for δ_0^* in Table 1 indicate that J.A. is sufficiently accurate for this energy beyond the second zero, $x_{02} = 0.0111$, of the wave function. However, the order, n , of the zero x_0 was determined by the amount of work involved in the numerical integrations rather than by the accuracy of J.A.

(b) High Order Phases

It is well known that for large l , J.A. is applicable for all x . The form used was

$$\delta_l = \int_{x_1}^{\infty} F_1^{\frac{1}{2}} dx - \int_{x_2}^{\infty} F_2^{\frac{1}{2}} dx,$$

where x_1 and x_2 are the zeros of F_1 and F_2 respectively. The differences, $\chi_l = \delta_l - \delta_{-l-1}$, were obtained from

$$\chi_l = \int_{x_1'}^{\infty} (F_1^{\frac{1}{2}})^{\frac{1}{2}} dx - \int_{x_1''}^{\infty} (F_1^{-l-1})^{\frac{1}{2}} dx,$$

where x_1', x_1'' are the zeros of F_1 for l and $-l-1$ respectively. In all cases, the integrals were evaluated graphically.

It should be noted that Table 2 shows that $|\delta_{\text{exact}} - \delta_{\text{J.A.}}|$ is small everywhere except for δ_0' . However, it is well known that the method breaks down for $l=0$. This agreement is much closer than that obtained by Bartlett and Welton.

* — is used for electrons, + for positrons.

TABLE 1
ACCURACY OF J.A.

x	δ_0^-	$\delta_l \left \frac{x_{n+1}}{x_n} \right $ by N.I.	$\delta_l \left \frac{x_{n+1}}{x_n} \right $ by J.A.
0.004	1.282	0.381	0.410
0.008	1.663	0.229	0.236
0.012	1.892	0.161	0.164
0.016	2.053	0.122	0.122
0.020	2.175	0.098	0.098
0.024	2.273	0.081	0.081
0.028	2.354	0.068	0.068
0.032	2.422	0.059	0.059
0.036	2.481	0.053	0.052
0.040	2.534		

TABLE 2
COMPARISON OF PHASES DETERMINED BY J.A. AND N.I.

Electrons					Positrons					
l	$\delta_l \left \frac{x_0}{0} \right $ by N.I.	$\delta_l \left \frac{x_0}{0} \right $ by J.A.	x_0	Order of Zero	$\delta_l \left \frac{x_0}{0} \right ^\infty$ by J.A.	$\delta_l \left \frac{x_0}{0} \right $ by N.I.	$\delta_l \left \frac{x_0}{0} \right $ by J.A.	x_0	Order of Zero	$\delta_l \left \frac{x_0}{0} \right ^\infty$ by J.A.
-4	0.825		0.015056	1	1.405	-0.820		0.019774	1	-1.24
-3	1.288		0.019375	2	1.251	-0.943		0.016961	1	-1.34
-2	1.789		0.014742	2	1.397	-1.165		0.014260	1	-1.42
-1	2.176		0.014194	2	1.408	-2.009		0.020730	2	-1.202
0	2.481	2.490	0.036000	5½	0.901	-2.059	-1.96	0.028709	3	-1.025
1	1.300	1.302	0.015997	2	1.341	-1.238	-1.238	0.022452	2	-1.156
2	1.087	1.09	0.019903	2	1.220	-1.045	-1.04	0.025438	2	-1.12
3	0.952	0.95	0.023506	2	1.120	-0.718	-0.71	0.019493	1	-1.26
4	0.646	0.64	0.018502	1	1.265	-0.649	-0.66	0.022348	1	-1.19
5	0.597	0.59	0.021524	1	1.195	-0.603	-0.59	0.025204	1	-1.12
6	0.557	0.55	0.024492	1	1.155	-0.567	-0.57	0.028037	1	-1.07
7	0.526	0.52	0.027408	1	1.105	-0.538	-0.55	0.030850	1	-1.04
9	0.482	0.48	0.033124	1	1.013	-0.489	-0.49	0.036418	1	-0.95

Table 3 gives the final values of the phases, all of which are estimated to have an error no greater than ± 0.01 radians.

TABLE 3
FINAL PHASE SHIFTS

Electrons				Positrons		
l	δ_l	$\chi_l = \delta_l - \delta_{l-1}$	δ_{l-1}	δ_l	$\chi_l = \delta_l - \delta_{l-1}$	δ_{l-1}
0	3.382*	(-0.202*)	(3.584*)	-3.084*	(0.127*)	(-3.211*)
1	2.641*	-0.55*	3.186*	-2.394*	0.19*	-2.584*
2	2.31*	-0.23*	2.54*	-2.16*	0.12*	-2.28*
3	2.07*	-0.16*	2.23*	-1.98*	0.08*	-2.06*
4	1.91*	-0.12	2.03	-1.84*	0.06	-1.90
5	1.79*	-0.09	1.88	-1.72*	0.04	-1.76
6	1.71*	-0.07	1.78	-1.64*	0.03	-1.67
7	1.63*	-0.06	1.69	-1.58*	0.03	-1.61
8	1.56	-0.05	1.61	-1.50	0.02	-1.52
9	1.50*	-0.04	1.54	-1.44*	0.02	-1.46
10	1.43	-0.04	1.47	-1.38	0.01	-1.39
11	1.37	-0.03	1.41	-1.32	0.01	-1.33
12	1.31	-0.03	1.34	-1.26	0.01	-1.27
13	1.26	-0.03	1.29	-1.21	0.01	-1.22
14	1.21	-0.02	1.23	-1.16		-1.16
15	1.16	-0.02	1.18	-1.12		-1.12
17	1.07		1.09	-1.04		-1.04
19	1.00	-0.01	1.01	-0.98		-0.98
21	0.94		0.95	-0.92		-0.92
23	0.89	-0.01	0.90	-0.88		-0.88
25	0.84		0.85	-0.84		-0.84
27	0.80		0.81	-0.80		-0.80
29	0.76		0.76	-0.76		-0.76
31	0.72		0.72	-0.72		-0.72
33	0.69		0.69	-0.70		-0.70
35	0.66		0.66	-0.67		-0.67
40	0.60		0.60	-0.62		-0.62
45	0.54		0.54	-0.56		-0.56
50	0.50		0.50	-0.52		-0.52
70	0.37		0.37	-0.40		-0.40
100	0.28		0.28	-0.31		-0.31
200	0.14		0.14	-0.18		-0.18

* Phases determined by numerical integration and Jeffrey's approximation.

IV. SUMMATION OF SERIES

The summation of the slowly converging Mott series for $f(0)$ was effected by a similar method to that used by Bartlett and Welton, who noticed that, for

large l , the phases fall off in an exponential manner. Thus, beyond $l=33$, it was found possible to fit the exponential functions

$$\left. \begin{aligned} A_{\infty}^{-}(l) &= 0.60e^{-0.0137l} + 8.36e^{-0.0795l} - 24.5e^{-0.15l} \\ A_{\infty}^{+}(l) &= 0.543e^{-0.0107l} + 3.75e^{-0.062l} - 5.0e^{-0.15l} \\ B_{\infty}^{-}(l) &= 1.00e^{-0.0065l} + 2.50e^{-0.06l} - 24.5e^{-0.15l} \\ B_{\infty}^{+}(l) &= -0.955e^{-0.005l} - 1.5e^{-0.05l} + 16.33e^{-0.15l} \end{aligned} \right\}$$

to the functions

$$A_l = \frac{l+1}{2l+1}(1 - \cos 2\delta_l) + \frac{1}{2l+1}(1 - \cos 2\delta_{-l-1})$$

$$B_l = \frac{l+1}{2l+1}(\sin 2\delta_l) + \frac{l}{2l+1}(\sin 2\delta_{-l-1}).$$

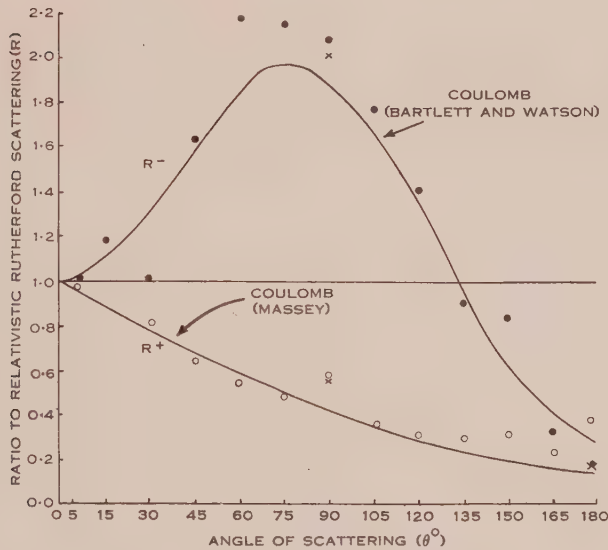


Fig. 1.—Results for angular distribution.

○ Calculated positron points.

● Calculated electron points.

The sum $f(\theta)$ can then be expressed as

$$\begin{aligned} \frac{2kf(\theta)}{i} &= \sum_{l=0}^{\infty} (2l+1)[A_{\infty}(l) - iB_{\infty}(l)]P_l(\cos \theta) \\ &+ \sum_{l=0}^8 (2l+1)[(A_l - A_{\infty}(l)) - i(B_l - B_{\infty}(l))]P_l(\cos \theta). \end{aligned}$$

The first term was summed analytically by making use of the formula

$$\sum_{l=0}^{\infty} (2l+1)e^{-\lambda l}P_l(\cos \theta) = \frac{1 - e^{-2\lambda}}{\{(1 - e^{-\lambda})^2 + 4e^{-\lambda} \sin^2 \theta/2\}}.$$

The second term was summed numerically up to $l=33$, after which it becomes zero.

Because of the rapid convergence of χ_l to zero, the series for $g(\theta)$ can also be summed numerically. Actually, $g(\theta)$ does not contribute much to $I(\theta)$, but is important in determining the asymmetry on double scattering. Thus, for the asymmetry 2δ on double scattering at 90° , viz.

$$2\delta = 2 |f \cdot g^* - f^* \cdot g|^2 / \{f^2 + g^2\}^2,$$

the values $2\delta^- = 0.03$, $2\delta^+ = 0.01$ were obtained. These check roughly with the corresponding Coulomb values $2\delta_c^- = 0.04$, $2\delta_c^+ = 0.008$.

The final results for the angular distribution (Fig. 1) are expressed as the ratio R of $I(\theta)$ to the relativistic Rutherford value, viz.

$$R = 4k^2/q^2 \cdot \sin^4 \theta / 2I(\theta),$$

where

$$q = 2\pi Ze^2/hv.$$

The crosses " \times " at 90° and 180° represent calculations of R using numerically summed values of $f(\theta)$. This is possible at these angles because of the properties of the $P_l(\cos 90^\circ)$ and $P_l(\cos 180^\circ)$. Thus we have an independent check on the methods of summation of $f(\theta)$.

V. DISCUSSION

For electrons, Figure 1 indicates that there are no large deviations from Bartlett and Watson's (1940) Coulomb values at large angles, as predicted by Mohr. One would expect, from simple considerations, that screening would only have an appreciable effect at low angles—which is the result obtained by Bartlett and Welton. However, at 1 MeV., such an effect does not appear to be present even at 5° . The increase in the cross section above Coulomb values in the region of 90° seems to be quite definite, but the deviations at 30° and 150° are not sufficient to confirm or deny the presence of previously suggested "diffraction" maxima and minima. It is unlikely, however, that these deviations are due to local errors, for the summations at 30° and 150° were checked independently. If the fault is in the phase shifts (say small errors in the high order phases), then the neighbouring points should also be slightly displaced.

Within the limits of accuracy of the calculation, the positron curve appears to be of Coulombian form for all angles except 130° – 180° . The apparent rise above Massey's (1942) values at these angles could be due to the greater sensitivity of the methods of summation to small errors in the phase shifts. At low angles, the sum does not depend so critically on such phase shift errors.

The value of R/R^c at 57.9° , viz. 3.3, agrees quite well with the recent experimental value of 3.70 obtained by Lipkin (1950) and Lipkin and White (1950) for platinum ($Z=78$) at 0.98 MeV. The corresponding Coulomb value is less than 3.0.

VI. ACKNOWLEDGMENTS

It is a pleasure to acknowledge the assistance of Associate-Professor C. B. O. Mohr and Mr. J. A. McDonell, who both showed continued interest in the problem.

VII. REFERENCES

- BARTLETT, J. H., JR., and WATSON, R. E. (1940).—*Proc. Amer. Acad. Arts Sci.* **74** : 53.
BARTLETT, J. H., JR., and WELTON, T. A. (1941).—*Phys. Rev.* **59** : 281.
FOX, L., and GOODWIN, E. T. (1949).—*Proc. Camb. Phil. Soc.* **45** : 373.
HARTREE, D. R., and HARTREE, W. (1935).—*Proc. Roy. Soc. A* **149** : 210.
LIPKIN, H. J. (1950).—Ph.D. Thesis, Princeton University, U.S.A.
LIPKIN, H. J., and WHITE, M. G. (1950).—*Phys. Rev.* **79** : 892.
MASSEY, H. S. W. (1942).—*Proc. Roy. Soc. A* **181** : 14.
MASSEY, H. S. W., and MOHR, C. B. O. (1941).—*Proc. Roy. Soc. A* **177** : 341.
MOHR, C. B. O. (1943).—*Proc. Roy. Soc. A* **182** : 189.
MOTT, N. F., and MASSEY, H. S. W. (1949).—"Theory of Atomic Collisions." 2nd Ed. (Clarendon Press : Oxford.)

THE DISTRIBUTION OF THE DISCRETE SOURCES OF COSMIC RADIO RADIATION

By B. Y. MILLS*

[*Manuscript received October 16, 1951*]

Summary

The galactic distribution of 77 discrete sources of cosmic radio radiation has been examined. It is found that the distribution can be explained on the assumption that there are two major classes of source, one having a high degree of galactic concentration and the other having a random distribution. The possible identification of the sources with various types of astronomical object is discussed, and it is concluded that, although some of the evidence is suggestive, it is insufficient for any positive conclusions to be formed.

I. INTRODUCTION

Although many research teams have now devoted considerable time to the study of the discrete sources of cosmic radio radiation or "radio stars", the main problems remain unsolved. The class or classes of objects which produce the radiation have not been definitely identified, and it is not known with certainty to what extent the integrated output of all the sources is responsible for the general background of radiation from the Galaxy.

Many ways of attacking the problem suggest themselves. Accurate position measurements and identification with unusual visual objects in the neighbourhood have been attempted by Bolton, Stanley, and Slee (1949), but this method must be treated with caution, as shown by the fact that one of the most powerful sources, that in Cygnus, has little possibility of such identification (Mills and Thomas 1951). Spectra have been measured (Ryle 1950; Stanley and Slee 1950) and a statistical analysis of a number of sources has been made (Ryle 1950). Most workers have also attempted to measure angular sizes, but until recently only upper limits have been obtained. While no single one of these methods can be expected to solve the main problems, it is probable that if they are all extended in their range a solution will be obtained.

The present observations have been made with the purpose of studying the galactic distribution of a large number of sources. Using a sensitive receiving system with a Michelson-type interferometer at a frequency of 101 Mc/s., a total of 77 sources has been observed. At this stage no attempt has been made to measure their positions with the full accuracy obtainable with the interferometer, but the accuracies are sufficient for statistical analysis.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

DISCRETE SOURCES OF COSMIC RADIO RADIATION



One of the three 101 Mc/s. aerials.

II. EQUIPMENT

A very similar type of interferometer has already been described (Mills and Thomas 1951); the present equipment differs only in the size, type, and number of aerials, and in the amplifying and recording method. It has been designed principally for accurate position measurements, but as accurate calibration of the whole equipment is rather a large undertaking, the distribution of sources was studied first. Three aerials are used, all on an east-west line, together with two receivers and two pen recorders so that the output from two

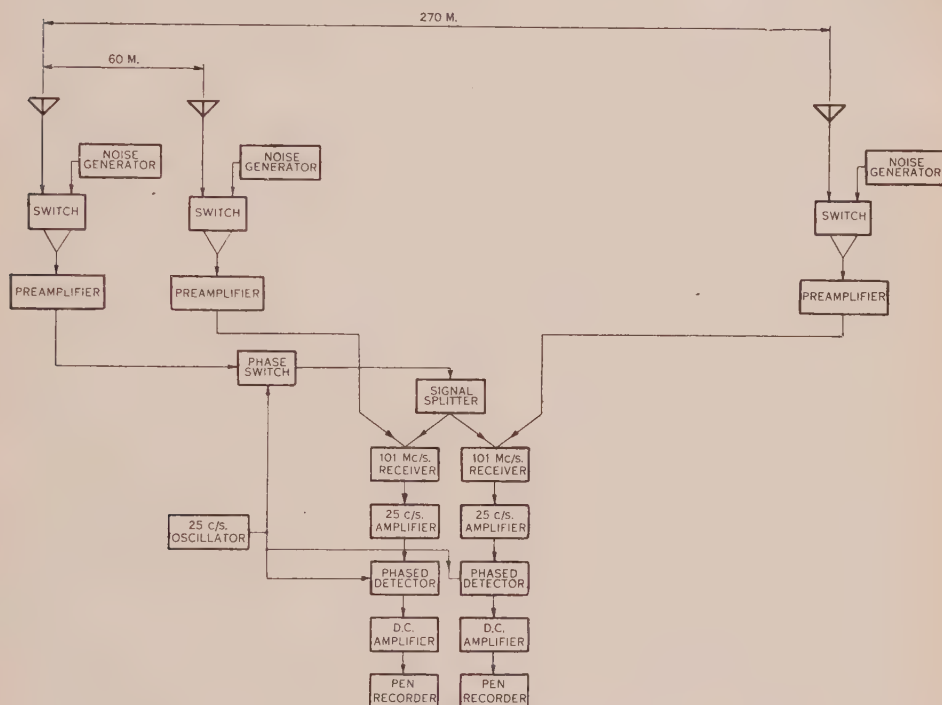


Fig. 1.—Block diagram of the equipment.

different aerial spacings may be recorded simultaneously. The maximum spacing is approximately 270 m., giving a 40' lobe spacing, and the minimum is approximately 60 m., giving a 3° lobe spacing. The general arrangement is shown in Figure 1 (from which the calibration system has been omitted) and a photograph of one of the aerials in Plate 1.

Each aerial consists of an array of 24 end-fed half-wave dipoles backed by a reflecting screen; it can be tilted about an east-west horizontal axis. The beam widths to the half-power points are approximately 24° in the east-west plane and 14° in the north-south plane. As before, preamplifiers are mounted close to the aerials, in this case in the box on the pyramid supports. All the cables between the aerials and the receiver hut are buried at a depth of 12 to 18 in. underground in an attempt to reduce fluctuations in electrical length due to temperature variations.

A major departure from the last equipment occurs in the amplifying and recording equipment, where it has been found that a method, the principle of which is due to Ryle,* has some advantages over the simple radio-frequency and low-frequency band-pass amplifiers used previously. Between one of the preamplifiers and the mixing point a switch is placed which can change the phase of the signal by 180° . The switch is of an electronic type and is actuated by a multivibrator arranged to have a switching rate of 25 c/s. The effect of this is to switch the interferometer lobe pattern backwards and forwards so that a maximum in the pattern is replaced by a minimum 25 times a second. A discrete source, if it is near a maximum or minimum in the pattern, will then cause a modulation of the amplitude of the mixed signal at a frequency of 25 c/s. After amplification and detection in the receiver system, the A.C.

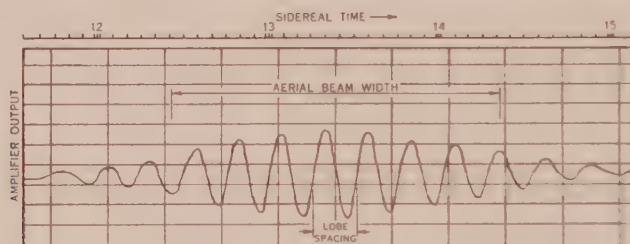


Fig. 2.—A typical pattern obtained with the smallest aerial spacing (source 13—4). With wider aerial spacings the envelope of the pattern is similar but the lobe spacing is smaller.

component of the rectified signal is fed to a phase-sensitive detector gated at the modulating frequency. The output of this second detector will be a D.C. signal which will be positive, zero, or negative, depending on the position of the source relative to the oscillating lobe pattern. The output from this detector is amplified and applied to a pen recorder. As a source passes in front of the aerials, the recorder therefore shows alternate excursions above and below the base line. It may be shown that, providing the power received from the source is small compared with the sum of the receiver and cosmic noise powers, the resulting pattern is sinusoidal in form, and its amplitude is proportional to the power received from the source. The period of the pattern will be the same as the normal pattern obtained without any switching, and may therefore be used to calculate the declination in the usual way. A typical pattern from a strong source (13—4) obtained with the smallest aerial spacing is shown in Figure 2. This system may be calibrated for a Right Ascension measurement in a similar way to that described by Mills and Thomas (1951) but, as explained above, it was uncalibrated for the work described here.

It has been found that this method of recording does not greatly increase the maximum sensitivity which may be obtained from the equipment, but it does allow a very considerable extension of the time for which the sensitivity is

* A full account of Ryle's work is being published in the *Proceedings of the Royal Society*.

available. This would appear to be due to two causes; firstly, a discrimination against forms of interference which are due to diffuse radiators, e.g. radiation from power lines, and secondly, a relaxation of the stability requirements of the radio-frequency amplifying equipment. Using a straight amplifying system, the best records were usually obtained between midnight and dawn; observations during the day were normally quite useless because of interference. Now it is found that the quality of the records does not change during the night, and in the absence of solar interference, good records are usually obtained during the day.

The simultaneous use of two aerial spacings has been useful for making rough estimates of the angular sizes of the sources. If the two receivers are adjusted for equal sensitivity, a point source will give a pattern of equal size on the two recorders. A source which has an angular extent in the east-west direction of $10'$ or more will produce a smaller pattern on the recorder fed by the aerials with the greater spacing, and from a knowledge of the ratio between the pattern sizes an equivalent angular width may be calculated.

III. OBSERVATIONS

Observations were made by setting all three aerials to a given declination, connecting the receivers to the aerials with maximum and minimum spacing, and allowing the equipment to run continuously for a period of at least 48 hours. If at the end of that period satisfactory records had been obtained, the aerial setting was changed by 10° and the procedure repeated. In this way the sky was gradually covered from a declination of $+50^\circ$ to -90° except for certain portions which culminated during the daylight hours when interference from the Sun was experienced. The whole procedure was then repeated at a later date to fill in these blank areas, which by then had moved into the night observing period. In actual practice many of these observations had to be repeated because of equipment trouble, electrical storms, and other interference. Altogether a total time of about 10 months from February to December 1950 was required for the observations. In every case at least two observations of every part of the sky (within the declination limits) have been made with the maximum aerial spacing, but the observations with the smallest spacing have not been so systematic, and some areas of the sky have not been covered at all.

No attempt has been made to identify individual lobes of the interference pattern, and the R.A. of a source is obtained by noting the sidereal time of the point of maximum interference pattern size, i.e. it is determined by the aerial radiation pattern alone. The declination is obtained by measuring the frequency of the interference pattern as described in the earlier paper (Mills and Thomas 1951). The accuracy is not very good, the stronger sources having an estimated probable error of 2 min. in R.A. and 0.15 per cent. in the cosine of the declination (corresponding to $20'$ of arc at a declination of 12°). The average probable error in the positions of the weakest sources is about 2° .

A certain difficulty arises if there is more than one source in the beam simultaneously, as beating between the superimposed patterns makes it difficult

to determine the point of maximum response for the various sources, or sometimes even to estimate the actual number of sources involved. The routine changes of declination by 10° steps are usually adequate for sorting out these beating effects. In a few places, however, they are not sufficient, and it has been found necessary to set the aerials on slightly unequal declinations. Since the magnitude of the source pattern is proportional to the geometric mean of the powers received by each aerial from the source, the effect of unequal declination settings is to give a narrower effective aerial beam. In this way most of the uncertainties were resolved.

In Figure 3 are shown two records of a weak source (01 + 3) taken on consecutive nights. The records have been superimposed to demonstrate the

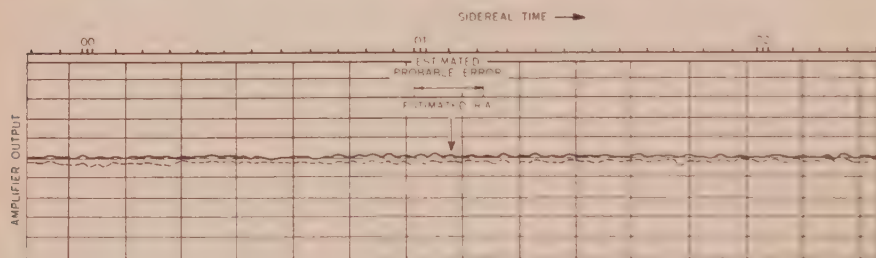


Fig. 3. Typical records of a weak source (01 + 3) obtained with the largest aerial spacing.
 ————— 9.ix.50. - - - - - 8.ix.50.

stability of the equipment. The estimated R.A. is shown, but it is clear that the small amount of beating which is present makes it difficult to estimate this with any accuracy.

Measurement of the intensity of a source involves the measurement of the interference pattern size and the effective temperature of the general background radiation. From a number of constants of the system which must be determined in addition, the incident power from the source may be calculated. These constants are the aerial gain, the preamplifier noise factor, and the total amplifier gain of the system. Since all these involve uncertainties, the final result is not very accurate. The probable error for the larger sources is estimated as about 10 per cent., rising to about 30 per cent. for the smaller ones.

IV. THE SOURCES

A list of the sources with their positions and intensities is given in Appendix I. For the purpose of analytical work on their distribution it has been found convenient to introduce a system of numbering the sources so that they may be readily found in any table or diagram. This involves attaching a number to each source which specifies its approximate position. First the hour of its R.A. is quoted, and then the sign and the tens digit of its declination. For instance, the source designated by Bolton as Cygnus-A becomes 19 + 4, Taurus-A becomes 05 + 2, and Centaurus-A becomes 13-4. If there is more than one source in the respective intervals, they are labelled A, B, C, etc., starting with the source

of lowest declination and proceeding to the highest. It is not suggested that the above system is superior to others or that it should be adopted generally. It has, however, proved very convenient for the work described in the present paper.

It has likewise proved desirable to have some logarithmic measure of the flux density of the source. Rather than use the antique and rather clumsy "magnitude" of visual astronomy, logarithms to the base 10 are used, corresponding to the use of the "bel" in radio engineering. A flux density of $10^{-25} \text{ W.m.}^{-2} (\text{c/s.})^{-1}$ has been arbitrarily taken as corresponding to a level of zero, and the level of the source is then given as $L = \log_{10} P \times 10^{25}$, where P is the flux density of the source measured in the same units as the reference flux density. All measurements of the flux density are assumed to apply to the power received in both planes of polarization, that is twice the power received with a plane-polarized aerial. It is suggested that a similar system could be adopted generally. If this were done, however, some thought should be given to the question of a zero level. Here the zero level has been chosen so that all the measured levels are small positive numbers, but with increasing equipment sensitivities negative levels will soon be obtained. Since it is desirable that all levels should have the same sign, it would therefore appear that there are two solutions. Either a very low level which is unlikely to be reached in the foreseeable future is taken as zero, or else a high level ((such as $1 \text{ W.m.}^{-2} (\text{c/s.})^{-1}$) is chosen so that all levels become negative, the numerical values increasing for weaker flux densities. In either case the frequency at which the level is measured must always be specified.

For the purpose of compiling the list of sources and the subsequent analysis, a discrete source has been defined in terms of a particular kind of pattern produced on the pen recorder. This pattern has a definite periodicity and length dictated by the declination, the aerial beam width, and the flux density. The R.A. of the source is obtained by noting the sidereal time of the centre of the pattern, the declination is obtained by measuring the period of the pattern and calculating in the usual way, and the flux density is obtained by measuring the amplitude of the pattern. The positions and intensities of the sources defined in this way, however, are likely to be dependent on the nature and dimensions of the receiving equipment, for the rather unsystematic observations with the smallest aerial spacing have shown that at least two sources (17-2B and 08-4), and probably also a third (13-4, Centaurus-A), give larger intensities with the closer spacing, that is with the coarser interference pattern. This suggests that these sources are either extended in an east-west direction or consist of two sources of very nearly equal intensity very close together. The latter possibility is unlikely for statistical reasons. In the analysis of the distribution of sources the results from the largest aerial spacing alone have been presented, because the unsystematic nature of the other observations could cause a weighting of the results. Actually the conclusions do not need alteration if the intensities given by the small aerial spacing are used.

Altogether 77 sources are listed. The position and intensity of the weaker sources (in fact even their existence in a few cases) is rather doubtful because

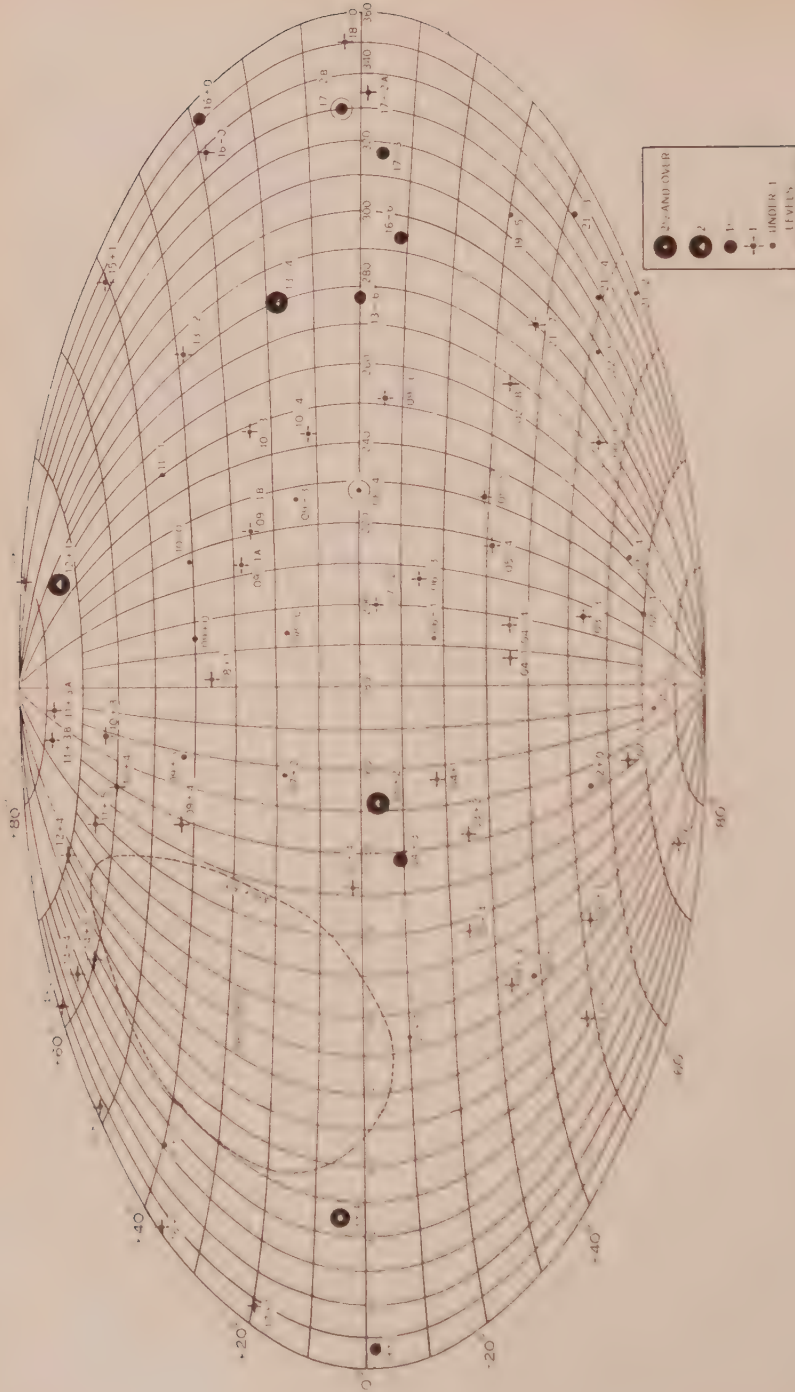


Fig. 4.—The galactic distribution of the discrete sources.

of the previously mentioned difficulty with beating between interference patterns which sometimes results in a pattern with more than one possible interpretation. Two such doubtful regions are in Taurus and in Fornax. It is noteworthy that there is some disagreement with Ryle (1950) and considerable disagreement with Stanley and Slee (1950) on the position and intensity of sources in the Taurus region. Since three different types of observation are involved, it is thought that the disagreements are due to instrumental effects.

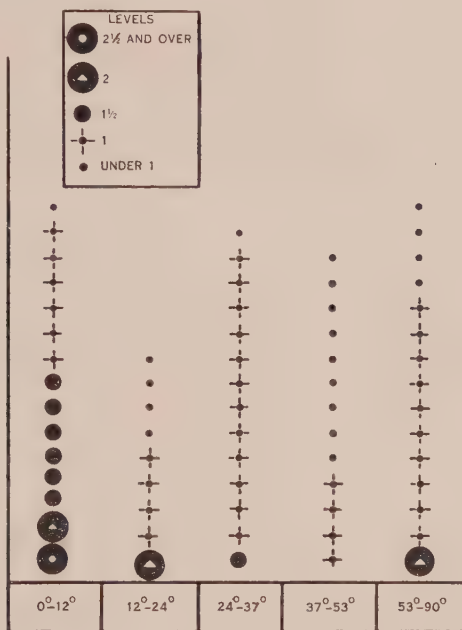


Fig. 5.—The concentration of the stronger sources towards the galactic plane.

The sources have been plotted on a galactic chart in Figure 4. The northern circumpolar region which is obscured is shown dotted. Also partially obscured is a region below and to the left of $19+4$ (Cygnus-A), because of the difficulty of making observations close to such a strong source. The sources with a definitely observable angular size already mentioned, 08-4 and 17-2B, are encircled by rings.

V. ANALYSIS

A qualitative estimate of the main features of the distribution may be made very simply. The sources shown in Figure 4 are collected together in Figure 5 in a form suitable for testing their galactic concentration. Ten equal-area bands are chosen, bounded by lines of constant galactic latitude, and in the figure the sources in each symmetrical pair of bands are arranged vertically. In order to overcome difficulties associated with the northern obscured area, all sources between $l=50^\circ$ and $l=130^\circ$ have been omitted. The most striking

feature of this diagram is the clustering of the more powerful sources towards the galactic plane. Other effects such as the scarcity of small sources in this region and their preponderance towards the galactic poles are thought to be instrumental effects, as will be explained later.

A more thorough analysis of the distribution may be made in the following way. It has been noted by Ryle (1950) and by Bolton and Westfold (1951) that with a homogeneous distribution of sources, all of the same absolute intensity, there is a definite relation between the number of sources observed and the sensitivity of the equipment. The relation is

$$N \propto P^{1.5},$$

where N is the number of sources received with a power P or greater. Piddington (unpublished data) and Westerhout and Oort (1951) have shown that the relationship holds also if the absolute source intensities are randomly distributed

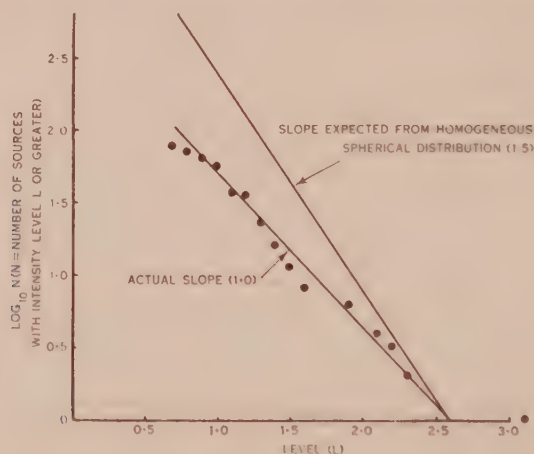


Fig. 6.—Number-level ogive of all the observed sources.

in various ways. This relation therefore provides a direct means of testing the homogeneity of any distribution. A similar relation has been used in analysing the distribution of external galaxies (Shapley 1943).

In Figure 6 the logarithm of the number of sources with a level L or higher is plotted against the level. It is seen that the slope of the ogive is approximately 1.0, as compared with a slope of 1.5 required by the above expression. A comparison with the visual stars shows that if their number-magnitude distribution is treated similarly, the initial slope is very nearly equal to 1.5 even though their distribution is not particularly homogeneous. The suggestion is, therefore, that the source distribution is very much less homogeneous than the stellar distribution.

In Figure 7, the galactic concentration is tested by separating the sources into two classes, those within $\pm 12^\circ$ of the galactic plane, and the remainder. Since the solid angles of the two areas are very different (1:1), it is necessary to plot the source density, i.e. the number of sources per steradian, instead of the

total number. Two conclusions can be drawn immediately: the density of the stronger sources is much higher in the vicinity of the galactic plane, and the slopes of the two curves are markedly different. The slope of the curve for the sources outside the galactic plane is consistent with a homogeneous distribution, while the slope for those in the vicinity of the plane is very much less. It may be observed that both slopes decrease rapidly at about the same source concentration. This is presumably due to insufficient resolution of the aerial beams, which causes the stronger sources to obscure weaker ones nearby. The absence of weak sources in the galactic plane, and their abundance at the galactic poles as demonstrated in Figure 5, is largely due to this effect, a contributing cause being the reduction in sensitivity of the equipment near the galactic plane, due to the increase in the general noise level.

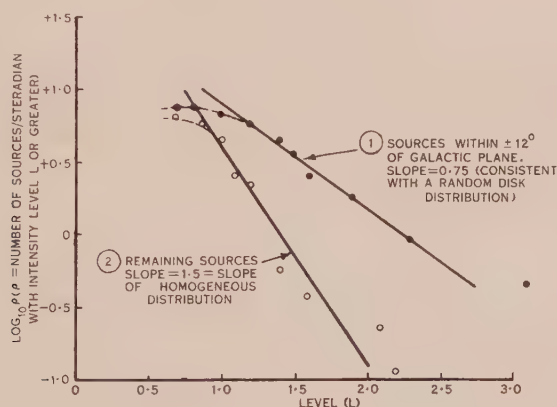


Fig. 7.—Comparison of sources: (a) within $\pm 12^\circ$ of the galactic plane (17 sources), (b) outside these limits (60 sources).

The small slope shown by the sources near the galactic plane is reasonable if a strong galactic concentration of a particular type of source is assumed, for it can be shown that the slope to be expected from a thin disk-shaped distribution is 1.0 if the observer is in the centre and a somewhat complicated result giving an average of less than 1.0 if the observer is near one edge. The observed slope of 0.75 is consistent with such a distribution.

It is interesting to observe the decreasing galactic concentration for the weaker sources, which is the opposite of the normal stellar distribution. From Figure 7 it can be seen that at an intensity level of 1.4, the galactic concentration ratio is 8 with a total of 16 sources. At an intensity level of 1.0, the ratio has decreased to 1.6 with a total of 55 sources.

As a check on the physical reality of the galactic concentration, a statistical analysis has been made to determine the possibility of such a concentration arising from an essentially random distribution. In Figure 8 the probability of the observed concentration, or a greater one, arising in this way is plotted against the intensity level. It can be seen that over most of the range the probability of such a chance concentration is extremely low, and therefore it is

reasonable to assume that the effect is a real one. The increase in probability towards the high levels is due to insufficient numbers of sources for a significant result, and the increase towards the low levels to the decreasing galactic concentration.

As Ryle (1950) has deduced a random distribution of sources from his observations in the northern hemisphere, it would be of interest to compare his results directly with the present ones. In Figure 18 of his paper is plotted an ogive similar to that of Figure 6 of the present paper, but the scale units are such that it is not easily related to the present approach. However, if it is replotted in a similar fashion to Figure 6, a slope of less than one is obtained, agreeing well with the present results. In a personal communication Ryle has

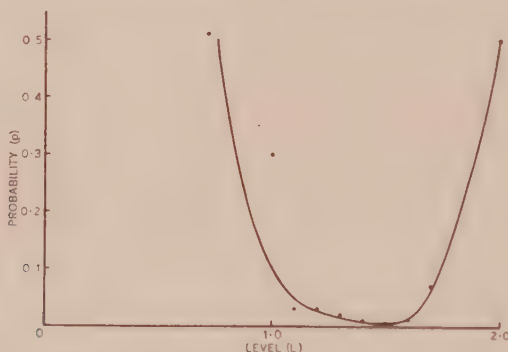


Fig. 8.—The probability, p , that the observed or a greater galactic concentration of sources with a level L or greater could arise from a random distribution.

informed the author that the unit of p , in his figure is incorrect, and that the intensities of some of the sources have subsequently been changed. It has therefore been considered inadvisable to reproduce the redrawn diagram. Further, since more than 90 per cent. of the sky has been covered in the present survey, the addition of the northern circumpolar sources is likely to have little effect on the observed distribution.

VI. DISCUSSION

It is reasonable to assume that two main statistical classes of source exist, for the graphs of Figure 7 and the diagram of Figure 5 indicate that there is one class which lies very close to the galactic plane (which we shall call class I), and another which is more or less randomly distributed (class II). Class I is naturally assumed to consist of powerful emitters distributed thinly through the Galaxy at great distances. Class II can be either even more powerful extra-galactic emitters or relatively feeble emitters at distances sufficiently short for their distribution to be unaffected by the galactic shape. The latter requirement would be fulfilled by a common type of star of low luminosity, but with a radio emission many orders of magnitude greater than the "quiet" Sun. Let us consider each of the two classes in turn.

(a) *The Class I Sources*

The class I sources are presumably at great galactic distances because of the high degree of galactic concentration and the small slope of the curve in Figure 7. The distance of the closest sources of this type may be estimated roughly in two ways.

The slope of the curve is the same as that of the equivalent stellar number-magnitude ogive for limiting magnitudes of about 15. Since the slope of this curve is a measure of the shape of the distribution, it is a reasonable assumption that the closest class I sources are at distances of the same order as the average distance of 15th magnitude stars, that is about 2500 parsecs. Also, by comparing Figure 4 with similar diagrams of the distribution of novae, a similarity may be noticed between the novae distribution and that of the stronger sources, whence it is probable that they are at similar distances. The nearest measured nova was at a distance of approximately 400 parsecs. The best estimate for the distance of the closest class I sources is therefore somewhere between these two distances, that is of the order of 1000 parsecs. It is perhaps significant that this is also approximately the distance of the Crab nebula, which has been tentatively identified with the source Taurus-A (05+2) (Bolton and Stanley 1949).

It is interesting to see if the general background radiation from the Galaxy could originate in the integrated output of the class I sources alone. Since the distance factor is doubtful, an order of magnitude comparison only is possible.

From Figure 7 it can be seen that the power, p , received from the n th source is given by

$$p = P_0 n^{-0.75},$$

where P_0 is the flux density of the closest "average" source obtained by noting the level at which the mean slope curve has an ordinate corresponding to one source. It is equal to $5.6 \times 10^{-23} \text{ W.m.}^{-2} (\text{c/s.})^{-1}$.

If an integration is performed over all the class I sources in the Galaxy (N), we obtain

$$P_{\text{(total)}} = P_0 \int_1^N n^{-0.75} dn = 4P_0 N^{1/4}.$$

The number N may be estimated from the distance to the closest source derived above if the assumption is made that the distribution is uniform throughout the Galaxy. If the power received with a beam width θ in the galactic plane is considered, we have approximately

$$P_\theta = 4P_0(R/a)^{3/2} \theta / 2\pi,$$

where R is the extent of the Galaxy in the chosen direction and a is the distance to the nearest source. In order to minimize errors due to the assumption of a uniform distribution throughout the Galaxy, let us calculate the power received from directions towards the galactic centre and anticentre. Towards the galactic centre we may take R as 20,000 parsecs. Taking the beam width θ

as approximately $\frac{1}{3}$ radian, we obtain, on inserting the previously estimated values for P_0 and a ,

$$P_0 = 5 \times 10^{-23} \text{ W.m.}^{-2} (\text{c/s.})^{-1}.$$

In the direction of the anticentre we find in a similar way on taking R as 5000 parsecs that the calculated value is

$$P = 2 \times 10^{-23} \text{ W.m.}^{-2} (\text{c.s.})^{-1}.$$

In calculating these values the sources of large angular size have not been included with their correct intensities because the analysis has been based entirely on the wide-spaced aerial results. If the corrected intensities from the close-spaced aerials are used, it is found that both these calculated values are doubled.

The total power actually received from these directions on the equipment described has been found to be $1.6 \times 10^{-21} \text{ W.m.}^{-2} (\text{c.s.})^{-1}$ in the direction of the galactic centre, and $5 \times 10^{-22} \text{ W.m.}^{-2} (\text{c.s.})^{-1}$ in the direction of the anti-centre. It would therefore appear unlikely that the integrated output of the class I sources alone is sufficient to account for the observed background radiation from the Galaxy.

It is also interesting to compare the radio emission from a typical class I source with that of the Sun. Taking the received power as $5.6 \times 10^{-23} \text{ W.m.}^{-2} (\text{c.s.})^{-1}$ as before, and its distance as 1000 parsecs, we find that the total radio emission at 100 Mc s. is some 10^{16} that of the "quiet" Sun.

There is some evidence as to the nature of these sources. The Taurus source (05-2) has been provisionally identified with the Crab nebula (Bolton and Stanley 1949), a gaseous nebula of some 6 minutes angular size. Also two sources (08-4 and 17-2B) show evidence of a minimum angular extent of about 35 minutes of arc. The source 05-2 appears to have a flat radio spectrum over the measured range of 60 to 1200 Mc s., suggesting strongly that the source is a thermally emitting thin gas.* Similarly the source 17-2B has apparently been observed by Piddington and Minnett (in press) at 1200 Mc s., with approximately the same intensity as we obtain at 101 Mc s. using the close-spaced aerials, also suggesting an origin in a thin gas. There is an anomaly here, however, in that Shain (unpublished data) finds that the intensity of what is probably the same source at 48 Mc s. is many times this value; thus the hypothesis of simple thermal radiation over the whole range of frequencies must be discounted. However, the evidence does seem to suggest that the class I sources could be nebulous objects of considerable angular size.

The Cygnus region provides both opposition and support to this theory. The large source in Cygnus (19-4) has an angular width of less than $1\frac{1}{2}$ minutes and a spectrum similar to the Virgo and Centaurus sources (12-4 and 13-4) (Stanley and Slee 1950) which in this analysis are considered as class II sources. It is of course possible that there is more than one type of source which shows a strong concentration in the galactic plane, in which case the sources 12-4 and

* Recent work by Stanley, Slee, and Bolton (unpublished data) suggests that the spectrum may not be as simple as this.

13-4 would probably be included with the 19+4 type on the basis of their similarity in spectra. Alternatively the 19+4 source could be considered as belonging to class II, its presence in the galactic plane being fortuitous. In support of the gaseous nebula theory, however, is the presence of a second source in Cygnus discovered by Piddington and Minnett (1951). They have been able to delineate the shape, which is found to consist of a strip 6° long and $1\frac{1}{2}^\circ$ wide. They have very tentatively identified it with some illuminated gas clouds in the vicinity.

Some observations of another powerful source near the galactic plane (Cassiopeia) have been made by Ryle and Smith (1948) and Ryle (1950). They set an upper limit to the angular size of this source of $6'$ of arc, and also find that it has a spectrum similar to Cygnus-A.* Thus, unless this source too can be classified with the class II sources, it would seem possible that the class I sources include at least two subclasses.

Westerhout and Oort (1951) have calculated the total integrated emission at 100 Mc/s. to be expected in various directions due to the ionized gas clouds associated with early-type stars. A rough calculation based on the temperature distribution given in Figure 1 of their paper suggests that with the aerial used in the present survey, a total power of 10^{-22} W.m.⁻² (c/s.)⁻¹ would be received from these gas clouds when pointed at the centre of the Galaxy. This is in agreement with the power estimated to be received from all the class I sources in this direction. As the gas clouds are all very close to the galactic plane, it therefore seems plausible that aggregations of gas clouds are responsible for some, at least, of the class I sources.

(b) *The Class II Sources*

Let us now consider the randomly distributed sources of class II. On the evidence presented here it has been impossible to differentiate between the two alternatives mentioned previously, that is the extra-galactic hypothesis and the nearby star hypothesis. Of course these hypotheses are not mutually exclusive.

It would be very convenient if the general background of galactic radiation could be ascribed to the sources after the manner of Ryle (1950), Bolton and Westfold (1951), and Westerhout and Oort (1951). However, there are certain difficulties associated with the spectra (Stanley, Slee, and Bolton, unpublished data) which show that this identification is far from being established. The evidence of the present distribution, as shown in Figure 5, at first glance suggests an extra-galactic origin, for the sources show a definite tendency to cluster towards the galactic poles in the manner of external galaxies. It has already been explained, however, that the effect is probably instrumental. Also the polar concentration of the external galaxies is known to be the result of obscuration towards the plane of the Galaxy. Since any such obscuration of the radio sources is unlikely, this particular similarity of distribution must be discounted.

* The spectrum which Ryle gives for Cygnus-A is, however, very different from that of Stanley and Slee (1950) or Stanley, Slee, and Bolton (unpublished data).

A direct test for an extra-galactic origin can be made, however, by examining the correlation between the positions of the class II sources and the presence of bright external galaxies. Bolton, Stanley, and Slee (1949) have already suggested that the sources Virgo-A (12+1) and Centaurus-A (13-4) may be identified with such galaxies. However, none of the survey positions are sufficiently accurate to attempt any such direct identification and resort must be had to statistical analysis. The analysis must be restricted to sources near the galactic poles to avoid the effects of obscuration.

Areas within 60° of the poles were considered, involving a total of 42 sources, and all nebulae listed in the Shapley-Ames catalogue which have positions within 2° of the source positions were noted. The angle of 2° was chosen as it is roughly the average limit of accuracy of the positions. The total number of such nebulae was 1.8 times the number which might be expected if there were a random relationship between the sources and the external galaxies. Since the position of some of the sources, notably Virgo-A (12+1) and Fornax-A (03-3), occurred in areas rich in galaxies this result is not very significant. A further check was made by counting the number of source positions which have at least one galaxy associated with them. This number was found to be not significantly different from the number to be expected from a random relationship.

These tests therefore indicate that there is no general correlation between the positions of the sources and the external galaxies.* However, the possibility that a rare type of galaxy is involved is not ruled out and is, in fact, slightly strengthened by a tendency for the positions of the stronger sources to occur in the direction of the clusters of galaxies. A further small piece of evidence is that the differences in the magnitudes and number of the external galaxies towards the north and south galactic poles are maintained by the sources. Summing up, it might be said that there is a suggestion, but no definite evidence, that some of the class II sources are extra-galactic.

The definite identification of a class of source with rare extra-galactic objects would be of great importance to cosmological theory. It is therefore interesting to see if the various quantities involved are of the correct order of magnitude. Let us assume that all the class II sources are extra-galactic, that Virgo-A (12+1) is a typical class II source, that it is connected with the Virgo cluster, and that the radiation received from the galactic poles at 100 Mc/s. is due solely to the integrated output of similar sources.† Proceeding in a similar manner to the analysis of the class I sources in Section VI (a) we find

* This would be expected from the work of Hanbury Brown and Hazard (1950), which shows that the intensities of normal extra galactic nebulae would be very much lower than the intensities of the class II sources.

† Observations by Stanley, Slee, and Bolton (unpublished data) of the polar spectrum suggest that not all the radiation could be due to sources, but the rise in output at the lower frequencies which they obtain is suggestive that at least a measurable proportion of it is due to this cause. Bolton has also suggested to the author that an accurate evaluation of the source power integral should take into account the effects of red shift: for an order of magnitude comparison, however, this is unnecessary.

that the "extent" of the universe is 10^9 parsecs. The value for the "distance" at which the velocity of recession of the galaxies equals the velocity of light is 6×10^8 parsecs (by extrapolation of observed velocities). Thus the assumptions made above lead to a reasonable result, but only further work can determine to what extent they are correct.

A study of the class II sources would not be complete without checking for a correlation with the rare galactic objects which show little concentration in the galactic plane, that is, planetary nebulae and galactic clusters. When a procedure similar to that used for the external galaxies was followed, no correlation could be found for the positions of either of these classes of object.

Finally, the possibility that all the sources are very numerous galactic objects and at close distances has been examined thoroughly by Bolton and Westfold (1951) and Westerhout and Oort (1951). The restriction of their arguments to the class II sources does not significantly alter their results, although some modification to their constants would be required to fit the data.

VII. ACKNOWLEDGMENTS

The author is indebted to Mr. A. B. Thomas, who was responsible for a large part of the early work involved in the setting-up and operation of the equipment, and to Mr. A. Watkinson for assistance in the construction and maintenance of the equipment and the taking of observations.

VIII. REFERENCES

- BOLTON, J. G., and STANLEY, G. J. (1949).—*Aust. J. Sci. Res. A* **2**: 139-48.
 BOLTON, J. G., STANLEY, G. J., and SLEE, O. B. (1949).—*Nature* **164**: 101.
 BOLTON, J. G., and WESTFOLD, K. C. (1951).—*Aust. J. Sci. Res. A* **4**: 476-88.
 HANBURY BROWN, R., and HAZARD, C. (1950).—*Nature* **166**: 901-2.
 MILLS, B. Y., and THOMAS, A. B. (1951).—*Aust. J. Sci. Res. A* **4**: 158-71.
 PIDDINGTON, J. H., and MINNETT, H. C. (1951).—*Aust. J. Sci. Res. A* **4**: 459-75.
 PIDDINGTON, J. H., and MINNETT, H. C. (1952).—*Aust. J. Sci. Res. A* **5**: 17.
 RYLE, M. (1950).—*Rep. Progr. Phys.* **13**: 184.
 RYLE, M., and SMITH, F. G. (1948).—*Nature* **162**: 462.
 SHAPLEY, H. (1943).—"Galaxies." p. 183 et seq. (The Blakiston Co.: Philadelphia.)
 STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res. A* **3**: 234-50.
 WESTERHOUT, G., and OORT, J. H. (1951).—*B.A.N.* **11**: 323.

APPENDIX I

A list of the discrete sources observed at a frequency of 101 Mc/s. with an aerial spacing of approximately 90 wavelengths. The estimated probable errors in the position measurements are indicated by the superscripts as follows:

Right Ascension	Declination
(1) ± 2 min.	(1) $\pm 5'$
(2) ± 4 min.	(2) $\pm 10'$
(3) ± 8 min.	(3) $\pm 20'$
(4) ± 16 min.	(4) $\pm 40'$
(5) ± 32 min.	(5) $\pm 1.5^\circ$

APPENDIX I (*Continued*)

Re- ference Number	Constel- lation	Position				Flux Density P (W.m. ⁻² (c/s.) ⁻¹ $\times 10^{-24}$) (both planes of polarization)	Level L $\log_{10} (P \times 10^{25})$	Remarks
		Equatorial Coordinates		Galactic Coordinates				
		R.A. (hr. min.)	Dec. (deg.)	l (deg.)	b (deg.)			
00-2	Pisces	00 55 ⁽⁴⁾	-24 ⁽⁵⁾	94	-38	0.5	0.7	Declined from beating patterns only; all measurements are doubtful
00-1	Pisces	00 55 ⁽³⁾	+11 ⁽⁴⁾	96	-51	1.5	1.2	
00-1	Cetus	00 00 ⁽²⁾	-14 ⁽⁴⁾	52	-74	1.5	1.2	
00-6	Tucana	00 05 ⁽²⁾	-61 ⁽³⁾	277	-56	1.5	1.2	
01-5	Cassiopeia	01 00 ⁽³⁾	+52 ⁽⁴⁾	93	-10	0.5	0.7	
01-3	Pisces	01 05 ⁽³⁾	+30 ⁽⁴⁾	96	-31	1.0	1.0	
01-2	Cetus	01 50 ⁽³⁾	-22 ⁽³⁾	163	-73	0.8	0.9	
01-4	Phoenix	01 40 ⁽³⁾	-49 ⁽⁴⁾	241	-65	0.8	0.9	Complicated beating patterns make the interpretation difficult
02-0	Cetus	02 35 ⁽³⁾	-4 ⁽³⁾	143	-54	0.5	0.7	
02-1	Cetus	02 00 ⁽²⁾	-11 ⁽⁴⁾	141	-65	1.0	1.0	
02-4	Phoenix	02 00 ⁽⁴⁾	-40 ⁽⁴⁾	220	-70	0.7	0.8	Complicated beating patterns make the interpretation difficult
02-8	Octans	02 20 ⁽³⁾	-85 ⁽²⁾	266	-33	1.5	1.2	
03-4	Perseus	03 10 ⁽²⁾	-42 ⁽³⁾	118	-12	2.4	1.4	
03-2	Taurus	03 45 ⁽²⁾	-21 ⁽⁴⁾	141	-25	1.0	1.0	
03-3	Fornax	03 20 ⁽²⁾	-37½ ⁽³⁾	206	-55	2.4	1.4	Probably Fornax A (Stanley and Slee 1959)

APPENDIX I (Continued)

Reference Number	Constellation	Position				Flux Density P (W.m. ⁻² (c/s.) ⁻¹ $\times 10^{-24}$) (both planes of polarization)	Level L $\log_{10} (P \times 10^{25})$	Remarks
		Equatorial Coordinates		Galactic Coordinates				
		R.A. (hr. min.)	Dec. (deg.)	l (deg.)	b (deg.)			
04+3	Perseus	04 30 ⁽³⁾	+31 ⁽⁴⁾	137	-9	3.0	1.5	Probably Taurus-C (Stanley and Slee 1950)
04+1	Orion	04 50 ⁽³⁾	+10 ⁽⁴⁾	157	-18	2.0	1.3	Hard to disentangle from 05+2; measurements are doubtful
04-1	Eridanus	04 40 ⁽²⁾	-18 ⁽⁴⁾	189	-36	1.5	1.2	
04-3	Columba	04 50 ⁽³⁾	-30 ⁽⁴⁾	198	-36	1.0	1.0	Complicated beating patterns make the interpretation difficult
05+4	Auriga	05 00 ⁽³⁾	+42 ⁽⁴⁾	133	+1	1.5	1.2	
05+2	Taurus	05 30 ⁽¹⁾	+22 ⁽³⁾	152	-4	19	2.3	Taurus-A (Stanley and Slee 1950)
05-4	Pictor	05 30 ⁽²⁾	-46 ⁽³⁾	219	-31	2.5	1.4	Probably Pictor-A (Stanley and Slee 1950)
05-5	Pictor	05 55 ⁽³⁾	-56 ⁽⁴⁾	231	-29	1.6	1.2	
06-1	Lepus	06 00 ⁽³⁾	-17 ⁽⁴⁾	191	-17	0.7	0.8	
06-3	Canis Major	06 30 ⁽⁴⁾	-30 ⁽⁴⁾	206	-15	1.0	1.0	Complicated beating patterns make the interpretation difficult
07+2	Gemini	07 15 ⁽²⁾	+27 ⁽⁴⁾	159	+19	0.8	0.9	
07-2	Canis Major	07 05 ⁽²⁾	-20 ⁽⁴⁾	200	-4	1.0	1.0	
08+1	Cancer	08 50 ⁽²⁾	+13 ⁽⁴⁾	183	+35	1.6	1.2	
08-0	Hydra	08 10 ⁽³⁾	-4 ⁽⁵⁾	195	+18	0.5	0.7	

APPENDIX I (*Continued*)

Reference Number	Constellation	Position				Flux Density P (W.m. ⁻² (c.s.) ⁻¹ $\times 10^{-24}$) (both planes of polarization)	Level L $\log_{10} (P \times 10^{25})$	Remarks
		Equatorial Coordinates		Galactic Coordinates				
		R.A. (hr. min.)	Dec. (deg.)	l (deg.)	b (deg.)			
08—4	Vela	08 35 ⁽²⁾	—42 ⁽³⁾	229	0	0.6	0.8	Possibly Puppis-A (Stanley and Slee 1950). Intensity with the closest spacing is 5, corresponding to an equivalent angular size of 33
09 + 4	Ursa Major	09 00 ⁽²⁾	+48 ⁽³⁾	139	+43	2.0	1.3	
09 + 3	Lynx	09 00 ⁽³⁾	+34 ⁽⁴⁾	159	+43	0.7	0.8	
09 + 0	Leo	09 25 ⁽³⁾	+8 ⁽⁵⁾	194	+40	0.8	0.9	
09 —1A	Hydra	09 20 ⁽¹⁾	—11 ⁽⁴⁾	212	—28	2.5	1.4	
09 —1B	Hydra	09 35 ⁽³⁾	19 ⁽⁴⁾	221	—25	1.5	1.2	
09 —3	Pyxis	09 20 ⁽³⁾	—30 ⁽⁴⁾	227	—15	0.8	0.9	Complicated beating patterns make the interpretation difficult
09 —6	Carina	09 55 ⁽⁴⁾	—62 ⁽³⁾	251	—6	1.0	1.0	Deduced from beating patterns only; all measurements are doubtful
10 + 4	Ursa Major	10 30 ⁽¹⁾	+44 ⁽³⁾	140	+60	2.2	1.3	
10 + 3	Leo Minor	10 35 ⁽²⁾	+35 ⁽⁴⁾	158	+63	1.0	1.0	
10 + 0	Sextans	10 05 ⁽³⁾	—5 ⁽⁵⁾	217	+41	0.5	0.7	Possibly Sextans A (Stanley and Slee 1950)
10 —3	Antlia	10 55 ⁽³⁾	—33 ⁽⁴⁾	247	—24	1.0	1.0	
10—4	Vela	10 10 ⁽²⁾	—42½ ⁽³⁾	242	—12	1.0	1.0	

APPENDIX I (Continued)

Re- ference Number	Constel- lation	Position				Flux Density P (W.m. ⁻² c/s.) ⁻¹ $\times 10^{-24}$) (both planes of polarization)	Level L $\log_{10} (P \times 10^{25})$	Remarks
		Equatorial Coordinates		Galactic Coordinates				
		R.A. (hr. min.)	Dec. (deg.)	l (deg.)	b (deg.)			
11+5	Ursa Major	11 40 ⁽²⁾	+50 ⁽³⁾	114	+65	2.0	1.3	
11+3A	Ursa Major	11 35 ⁽²⁾	+31 ⁽⁴⁾	165	+76	1.0	1.0	
11+3B	Ursa Major	11 45 ⁽³⁾	+37 ⁽⁴⁾	140	+76	1.0	1.0	
11-1	Crater	11 45 ⁽³⁾	-14 ⁽⁴⁾	250	+46	0.5	0.7	Deduced from beating patterns only
12+4	Canes Venatici	12 55 ⁽³⁾	+49 ⁽⁴⁾	83	+70	1.2	1.1	
12+1	Virgo	12 30 ⁽¹⁾	+12 $\frac{1}{2}$ ⁽³⁾	260	+74	12	2.1	Virgo-A (Stanley and Slee 1950)
13+2	Coma Berenices	13 15 ⁽²⁾	+25 ⁽⁴⁾	340	+81	1.5	1.2	
13-2	Virgo	13 20 ⁽³⁾	-22 ⁽⁵⁾	282	+39	1.0	1.0	Pattern beats with 13-4; measure- ments are doubtful
13-4	Centaurus	13 20 ⁽¹⁾	-43 ⁽²⁾	278	+18	16	2.2	Centaurus-A (Stanley and Slee 1950). In- tensity with the closest spacing is 24; a very rough esti- mate of the equi- valent angular size is 20'
13-6	Centaurus	13 35 ⁽³⁾	-60 $\frac{1}{4}$ ⁽¹⁾	276	0	7.5	1.9	Beats with 16-6
14+5	Bootes	14 20 ⁽³⁾	+51 ⁽⁴⁾	57	+60	2.0	1.3	
14+4	Bootes	14 40 ⁽³⁾	+42 ⁽⁴⁾	38	+62	1.0	1.0	
14+2	Bootes	14 40 ⁽²⁾	+26 ⁽⁴⁾	3	+63	1.0	1.0	
15+2	Serpens Caput	15 30 ⁽³⁾	+26 ⁽⁴⁾	7	+52	1.0	1.0	

APPENDIX I (Continued)

Re- ference Number	Constel- lation	Position				Flux Density P (W.m. ⁻² (c.s.) ⁻¹ $\times 10^{-24}$) * (both planes of polarization)	Level L $\log_{10} (P \times 10^{25})$	Remarks
		Equatorial Coordinates		Galactic Coordinates				
		R.A. (hr. min.)	Dec. (deg.)	l (deg.)	b (deg.)			
15+1	Serpens Caput	15 10 ⁽²⁾	+11 ⁽⁵⁾	342	+51	1.0	1.0	
16+4	Hercules	16 40 ⁽²⁾	+41 ⁽⁴⁾	31	+48	0.8	0.9	
16+1	Hercules	16 30 ⁽²⁾	+18 ⁽⁴⁾	2	+36	1.3	1.1	
16+0	Hercules	16 45 ⁽¹⁾	+6 ⁽⁵⁾	351	+28	4.0	1.6	Hercules-A (Stanley and Slee 1950)
16-0	Ophiuchus	16 15 ⁽²⁾	-5 ⁽⁵⁾	337	+29	2.0	1.3	
16-6	Triangulum Australis	16 10 ⁽³⁾	-60 $\frac{3}{4}$ ⁽¹⁾	293	-9	8.5	1.9	Beats with 13-6
17-1	Ophiuchus	17 40 ⁽²⁾	+14 ⁽⁴⁾	5	+20	1.5	1.2	
17-2A	Sagittarius	17 55 ⁽³⁾	-23 ⁽⁴⁾	334	-1	1.5	1.2	Hard to disentangle from 17-2B
17-2B	Sagittarius	17 55 ⁽²⁾	-29 ⁽³⁾	330	+4	3.0	1.5	Intensity with closest aerial spacing is 30, corresponding to an equivalent angular size of 35'
17-3	Scorpius	17 20 ⁽²⁾	-39 ⁽³⁾	317	-4	4.0	1.6	
18-0	Ophiuchus	18 10 ⁽²⁾	-6 ⁽²⁾	351	+4	2.5	1.4	Possibly Serpens Cauda-B (Stanley and Slee 1950)
19+4	Cygnus	19 57 $\frac{1}{2}$	+40 $\frac{1}{2}$	44	+5	130	3.1	Cygnus-A (Mills and Thomas 1951)
19+0	Aquila	19 00 ⁽³⁾	+7 ⁽⁵⁾	7	-2	3.0	1.5	Deduced from beating pattern only (beats with 19+4); all measurements are doubtful

*Recent measurements have suggested the possibility that the extended source is not identical with 17-2B.

APPENDIX I (Continued)

Re- ference Number	Constel- lation	Position				Flux Density P (W.m. ⁻² (c/s.) ⁻¹ $\times 10^{-24}$) (both planes of polarization)	Level L $\log_{10} (P \times 10^{25})$	Remarks
		Equatorial Coordinates		Galactic Coordinates				
		R.A. (hr. min.)	Dec. (deg.)	l (deg.)	b (deg.)			
19—5	Telescopium	19 40 ⁽³⁾	—50 ⁽⁴⁾	316	—30	0.5	0.7	
20—0	Aquila	20 20 ⁽⁴⁾	—6 ⁽⁵⁾	8	—24	1.0	1.0	Could not be observed directly because of interference from 19—4; all measure- ments are doubtful
21—2	Capricornus	21 55 ⁽²⁾	—24 ⁽⁴⁾	356	—53	0.8	0.9	
21—3	Micro- scopium	21 00 ⁽³⁾	—31 ⁽⁴⁾	341	—43	0.8	0.9	
21—4	Crus	21 25 ⁽³⁾	—41 ⁽⁴⁾	328	—48	0.5	0.7	
21—7	Pavo	21 00 ⁽⁴⁾	—71 ⁽³⁾	290	—37	2.0	1.3	
22—5	Indus	22 00 ⁽⁴⁾	—54 ⁽⁴⁾	308	—51	0.5	0.7	
23+1	Pegasus	23 35 ⁽³⁾	+10 ⁽⁵⁾	64	—48	0.9	1.0	Hard to disentangle from 19+4

ON THE HYSTERESIS OF ADSORPTION ON SOLID SURFACES

By R. G. WYLIE*

[*Manuscript received December 24, 1951*]

Summary

Hysteresis phenomena associated with the adsorption of gases on solid surfaces are usually explained in terms of three-dimensional capillary effects or with more or less unspecific reference to phase transitions.

It is shown that hysteresis effects are to be expected when two-dimensional phase transitions occur on solids. In this connection, the thermodynamic equation governing the equilibrium of small, incompressible two-dimensional phases is derived. Such phases can form on an imperfect solid surface in an irreversible manner and, as calculation shows, can contribute significantly to the hysteresis of adsorption. In some cases the phase change may be responsible for the whole effect.

The diffuseness of first-order phase transitions may be due to the same mechanism.

I. INTRODUCTION

As is well known, the adsorption of gases at solid surfaces is commonly associated with hysteresis effects. These effects have been found not only with gels and substances whose surfaces are known to be porous but with surfaces not ordinarily associated with porosity such as mica (Bangham and Mosallam 1938; Pierce and Smith 1950). Existing theories of adsorption hysteresis fall broadly into two classes, namely those depending on capillary condensation, following the ideas of Zsigmondy (1911), and those relating in a more or less unspecified manner to phase transitions.

The capillary-condensation theory has been subjected to analysis by Katz (1949) who, testing the experimental data with derived self-consistency rules, concluded that the theory cannot be generally valid. The considerations of Pierce and Smith also involve the idea of three-dimensional condensates and these writers have suggested that the coalescence of patches of such condensates provides a mechanism with hysteresis. Gregg (1949) has suggested that it is natural to associate hysteresis with phase transitions, since such an association frequently occurs in bulk matter and since it appears that the time required to achieve equilibrium is considerably greater for pressures at which a transformation occurs.

In the present paper it is shown that hysteresis can be explained in terms of the equilibrium properties of two-dimensional phases, without reference to capillaries or to visco-elastic effects in the adsorbate. The theory is partly analogous to that for phase-embryo formation in three dimensions (Volmer 1939) but important differences, associated with the two-dimensional character and with imperfections of the crystal surface, exist.

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

The occurrence of phase transitions in films on solid surfaces, anticipated by Frenkel as long ago as 1924, has now been established (Gregg and Maggs 1948 ; Gregg 1949). Some, but not many, of these transitions are relatively sharp transitions of the first order but the commonest type appears to be the diffuse transition of first order. The theory which follows assumes that two surface phases in equilibrium are separated by a line boundary with which a line free energy is associated. This assumption represents an extension of the idea of linear free energy, for some time associated with the theory of crystal growth, and recently employed with remarkable success by Frank (1949) and Burton, Cabrera, and Frank (1951) in developing the theory of growth of real crystals. It should, perhaps, be pointed out that two surface phases can be in equilibrium with one another without being in contact, for the gas over the surface (or solute, in the case of a solution) has the same thermodynamic potential as the adsorbate in both surface phases. However, this fact is probably of no great significance, although the role of the three-dimensional phase cannot be ignored altogether.

Given the occurrence of a phase transition on the surface, many possible mechanisms can be suggested to account for hysteresis. The conditions which must be satisfied for these to operate are probably fulfilled in practice in many cases. The theory which follows may, with some justification, be regarded as a treatment of two-dimensional "capillary" effects. Surface defects of types which are probably extremely common in practice play a role in some ways analogous to that played in three dimensions by capillaries in the ordinary sense. The theory suggests that the transitions will take place in just such a manner as to give rise to hysteresis and so that the observed phenomena, whilst they may relate to essentially sharp fundamental processes, may be of a diffuse character.

II. THE EQUILIBRIUM OF SMALL PHASES IN TWO DIMENSIONS

It is convenient first to derive the conditions for equilibrium of two surface phases one of which is small, and the distinguishing boundary line between which is associated with a free energy, σ , per unit length. It will be assumed that σ is isotropic, that is, that the value of σ is independent of the direction of the boundary line on the substrate. This will not generally be so when the substrate is crystalline nor when the substrate is amorphous and the two-dimensional phase which is forming is of crystal-like character. However, the assumption does not seriously limit the validity of the arguments which follow.

Consider two surface phases of areas A and B which together occupy the whole constant area of an adsorbent. Let the adsorbent be in contact with a three-dimensional gas of constant pressure, p , with which the surface phases are in equilibrium. Let the molecular Helmholtz free energies in the surface phases be ψ_α and ψ_β and the respective molecular areas α and β . The molecular Helmholtz free energy and molecular volume of the gas will be denoted by ψ and v . The phase of lower energy, of area A , and associated with the subscript α , will be assumed to be of small extent and of circular shape with radius r .

The numbers of molecules in the surface phases are

$$g_{\alpha}=A/\alpha \text{ and } g_{\beta}=B/\beta.$$

The Gibbs free energy of the whole system, which contains N molecules in all, is

$$G=(N-g_{\alpha}-g_{\beta})\psi+g_{\alpha}\psi_{\alpha}+g_{\beta}\psi_{\beta}+p(N-g_{\alpha}-g_{\beta})v+2\pi r\sigma. \quad (1)$$

Noting that the lowering of surface free energy due to the β -phase, Π , is given by

$$\Pi=-\frac{\partial\psi_{\beta}}{\partial\beta}, \quad (2)$$

and that

$$dA+dB=0,$$

and assuming that the α -phase is incompressible, the equilibrium conditions

$$\frac{\partial G}{\partial g_{\alpha}}=0 \text{ and } \frac{\partial G}{\partial g_{\beta}}=0 \text{ are found to give}$$

$$-\psi+\psi_{\alpha}+g_{\beta}\frac{\partial\psi_{\beta}}{\partial g_{\alpha}}-pv+2\pi\sigma\frac{\partial r}{\partial g_{\alpha}}=0,$$

and

$$-\psi+\psi_{\beta}+g_{\beta}\frac{\partial\psi_{\beta}}{\partial g_{\beta}}-pv=0,$$

which reduce to

$$(\psi+pv)-(\psi_{\alpha}+\Pi\alpha)=\frac{\sigma}{r}\alpha, \quad (3)$$

and

$$(\psi_{\beta}+\Pi\beta)-(\psi+pv). \quad (4)$$

Equations (2) and (4) give the Gibbs adsorption isotherm in the form

$$d\Pi=\beta^{-1}d\mu, \quad (5)$$

where μ is written for $(\psi+pv)$ as is usual. When the phase α is large, (3) gives

$$\mu_0-(\psi_{\alpha}+\Pi_0\alpha)=0,$$

so that (3) may be written

$$(\mu-\mu_0)-(\Pi-\Pi_0)\alpha=\frac{\sigma}{r}\alpha. \quad (6)$$

Using (5) and the explicit expression for the thermodynamic potential of a perfect gas, the last equation becomes

$$kT\left(\log\frac{p}{p_0}-\alpha\int_{p_0}^p\beta^{-1}d\log p\right)=\frac{\sigma}{r}\alpha. \quad (7)$$

When α is small compared with β , as in the condensation of a gaseous film to form a compact two-dimensional phase, the integral term in (7) may be neglected. The two-dimensional analogue of Thomson's equation is then obtained

$$kT\log\frac{p}{p_0}=\frac{\sigma}{r}\alpha. \quad (8)$$

When α is comparable with β , as is very often the case for phase transitions in two dimensions, the form (7) must be retained and, if the compressibility of the α -phase is considerable, the argument must be modified to give a still more general result. Equation (8) has been given by Burton, Cabrera, and Frank (1951) in connection with a study of the properties of two-dimensional embryos on substrates of the same substance. These authors have developed a generalized two-dimensional Wulff theorem, and considered the equilibrium properties when σ is not isotropic. For present purposes the α -phase may be treated as circular in shape.

The interpretation of equation (8) is that a phase equilibrium, which for large surface phases α and β exists at pressure p_0 , exists at a pressure p when the α -phase is small, of radius r . The concepts of the excess internal surface-pressure of the α -phase and of the linear tension of the boundary line have not

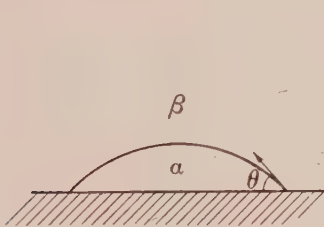


Fig. 1 (a)

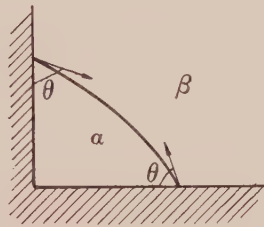


Fig. 1 (b)

been introduced since they are not of general significance. If the α -phase is bounded in part by solid obstructions then, depending on the relative magnitudes of the line free energies, contact angle phenomena will occur. Equation (8) remains valid but the possibility of the curvature of the circular boundary arcs being convex towards the phase α , corresponding to $p < p_0$, now exists.

The simplest case is that of an α -phase at a step. Figure 1 (a) shows an α -embryo bounded by a circular arc and a step. The shaded area represents a part of the substrate surface, parallel to the plane of the paper, which is at least one molecular layer higher than the unshaded area on which the β -phase has grown. The expression for the free energy of the whole system in this case resembles (1) except that the linear term, $2\pi r\sigma$, is replaced by $2r(\sigma\theta + \sigma' \sin \theta)$, where σ' is the excess of the step- α line free energy over the step- β line free energy and θ is the contact angle, which may be regarded as a further independent variable. The relationship between g_α and r is now

$$g_\alpha = r^2(\theta - \frac{1}{2} \sin 2\theta)/\alpha.$$

The equilibrium conditions $\partial G/\partial g_\alpha = 0$ and $\partial G/\partial \theta = 0$ are easily found to reduce to two equations which, taken together, are equivalent to equation (7) and a two-dimensional analogue of Neumann's equation, namely

$$\sigma \cos \theta + \sigma' = 0. \quad (9)$$

This equation defines the contact angle in terms of the line free energies. The free energy of formation of the equilibrium embryo, analogous to the result of Volmer (1929) for a part-sphere on a solid surface, is found to be

$$\Delta G = \sigma r (\theta - \frac{1}{2} \sin 2\theta). \quad \dots\dots\dots (10)$$

The work of formation thus increases linearly with r . It is worth noting that, for any configuration which is compatible with the embryo preserving its equilibrium shape with increasing size, the work of formation is one-half the total line free energy, or, what amounts to the same thing, the negative of the area free energy, for with preservation of shape,

$$\Delta G = ar^2 + br,$$

where a and b are functions of θ only, and the equilibrium condition is $r_0 = -b/2a$. Then $\Delta G = (-b/2 + b)r_0$, where the two terms arise from the area and line free energies respectively.

The case one degree more complex, but one for which shape is still preserved, is that of an embryo of α -phase forming at the intersection of two steps (Fig. 1 (b)). Assuming that the steps intersect at right angles, the free energy of formation of the equilibrium embryo is found to be

$$\Delta G = \sigma r [\theta - \pi/4 - \cos \theta (\sin \theta - \cos \theta)], \quad \dots\dots\dots (11)$$

ΔG , of course, increasing linearly with r .

The nucleation of phase α in either of these forms requires a free-energy fluctuation ΔG and, employing the fluctuation theory of Einstein (1910), as is conventional in the theory of three-dimensional nucleation, the nucleation probability per unit time, J , will vary with ΔG and temperature according to

$$J = Ce^{-\Delta G/kT}, \quad \dots\dots\dots (12)$$

where C is a constant. It is reasonable to assume that C will be of the same order of magnitude for all those nucleation processes which must begin at specific points and again of a common order of magnitude per unit length for all those processes the probabilities of which are uniform along a line. It may happen, of course, that nucleation at the intersection of steps occurs very easily: indeed, if the contact angle is less than $\pi/4$, nucleation will occur there for a negligible supersaturation.

The necessity of supplying the free energy of activation, ΔG , already provides one means of explaining hysteresis effects. As the vapour pressure (or surface pressure, Π) is increased, a value p , greater than the equilibrium pressure for the co-existence of α - and β -phases, may be reached before the transition begins. On the other hand, the reverse process, of "evaporating" the α -phase, goes to completion for p only infinitesimally smaller than p_0 . This mechanism, however, is not the only one which may give rise to hysteresis. A prominent part may be played by embryos of changing shape which will now be discussed.

When $\theta > \pi/2$ the complete locus of the centre is $AS'T'$ to infinity. (The two branches SRT and $R'S'T'$ of the same hyperbola correspond to supplementary angles of contact.) There is no range of stable equilibrium in this case. When $\theta < \pi/4$ a negligible activation energy is necessary to nucleate formation of the α -phase and the radius at first increases monotonically with increasing size. The centre follows the line AS' and then the branch $S'T'$ of a hyperbola to infinity. It then returns from infinity along the curve SR and traces out the whole curve

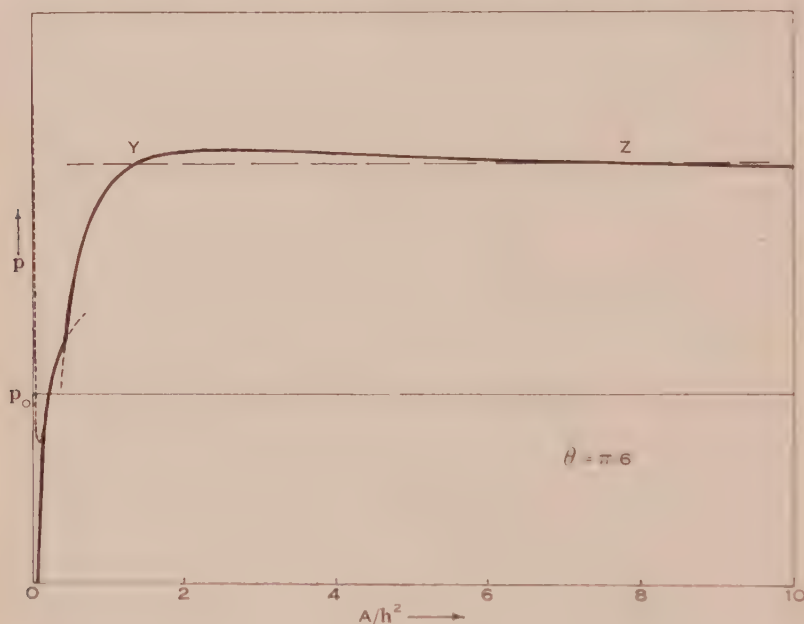


Fig. 3

RT , again going to infinity. In the extreme case $\theta = 0$, the radius goes monotonically from zero to infinity as growth occurs, the curvature of the boundary line remaining convex towards the α -phase. The locus is $AS'T'$ and the line AS' is a tangent of the conic which is now a parabola.

In this case a hysteresis effect is obtained even when $\theta = \pi/4$ and nucleation is easy, for the pressure must exceed the value corresponding to radius RP before the embryo can grow indefinitely. Figure 3 shows the equilibrium pressure plotted against area for $\theta = \pi/6$. The maximum p corresponds to radius RP , that is to $r = h/(1 - \cos \theta)$. In the evaporation process the representative point moves along the A -axis until it meets the p - A curve. The embryo then vanishes in equilibrium with the decreasing pressure as the point moves along this curve. The area for radius RP is

$$A = \frac{h^2(\theta - \frac{1}{2} \sin 2\theta)}{2(1 - \cos \theta)^2} \dots \dots \dots (16)$$

If the angles between intersecting steps are rounded off, as may be the case when the substrate has been partially annealed, an activation energy of nucleation

exists for any contact angle. The curve of Figure 3 is then modified as shown by the dotted line. In addition to the question of the nucleation of the first elements of α -phase at the point A (Fig. 2), another question of temperature-activated nucleation can be asked, namely, given that stable equilibrium is established at the point Y of Figure 3, what is the probability that the embryo

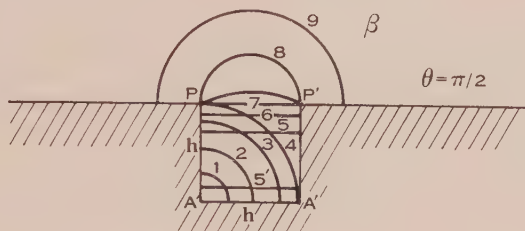


Fig. 4 (a)

will pass spontaneously over the maximum, past the point Z , and hence grow to visible size? In other words, as the pressure is raised, a value less than the maximum of Figure 3 will be reached at which the embryo will grow rapidly to visible size. However, unless the embryo is particularly small, this effect may be neglected, as it will be in what follows. When $\theta < \pi/4$ stable elements of α -phase

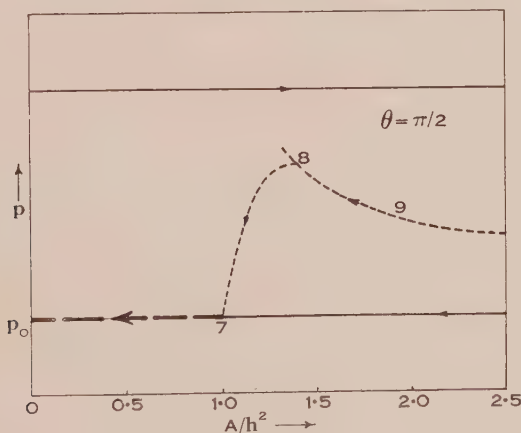


Fig. 4 (b)

can be retained at rectangular corners when the pressure is reduced appreciably below p_0 . The analogous phenomenon in three dimensions has been discussed by Turnbull (1950) following the ideas of Volmer (1939).

It may now be shown that, when the model is only slightly more complicated, essentially discontinuous change of shape can occur. The possibilities are so numerous that only a few examples can be given. Consider a step with the profile shown in Figure 4 (a), the contact angle being $\pi/2$. Nucleation at the point A may be assumed to require a vapour pressure higher than any which can be an equilibrium pressure for the particular model. Thus, when nucleation occurs, the embryo passes rapidly through the stages labelled 1, 2, 3 . . . 9,

to a large size. The evaporation of the α -phase, however, proceeds to completion for p infinitesimally less than p_0 , passing through the stages 9, 8, 7, 6, 5'. The actual vapour pressure, assumed to be changing very slowly, is plotted against the area in Figure 4 (b). The line horizontal straight line represents the rapid growth to large size for the constant pressure of nucleation. During evaporation,

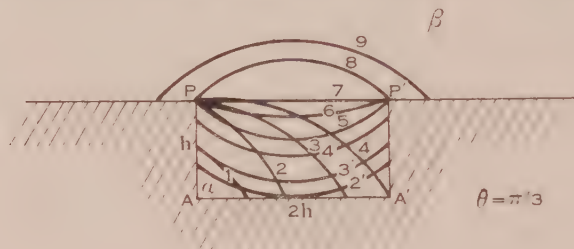


Fig. 5 (a)

the α -phase follows the fine line on the axis until the point labelled 7 is reached (a point corresponding to profile 7 of Figure 4 (a)). From this point the α -phase will evaporate in equilibrium following the heavy broken line. The fine broken line

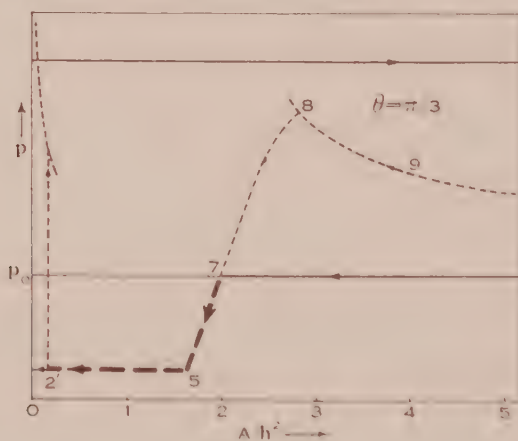


Fig. 5 (b)

line shows the equilibrium pressure corresponding to the actual forms which the α phase takes during the first stage of evaporation. The case of Figures 4 (a) and 4 (b) has no hysteresis other than that due to nucleation at the point A. Incidentally, the probability that nucleation will occur simultaneously at A and A' is entirely negligible and once it has occurred at A the rapid removal of β -phase (or three-dimensional gas) to form the α -phase will cause a drop in thermodynamic potential at A' which will prevent nucleation at that point during the growth of the α -embryo.

Figures 5 (a) and 5 (b) represent the growth at a step of partly rectangular profile, the contact angle being $\pi/3$. The growth process resembles that of Figure 4 (a) and is again represented simply by the fine horizontal line. However,

during evaporation the representative point follows the A -axis as far as the point 7 (fine line signifying that equilibrium does not obtain), then follows the heavy broken line (with equilibrium). The area decreases at constant pressure when the profile leaves the points P and P' and until it touches the line AA' . The phase then vanishes at constant pressure as signified by the short fine line

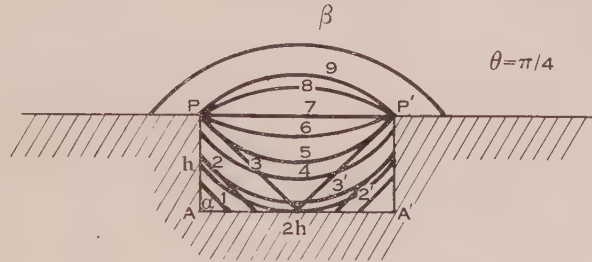


Fig. 6 (a)

of Figure 5 (b). The fine dotted curves again give the corresponding equilibrium pressures for the ranges where equilibrium does not actually obtain. The area of the hysteresis loop is observed to include a part which is not attributable to the nucleation process.

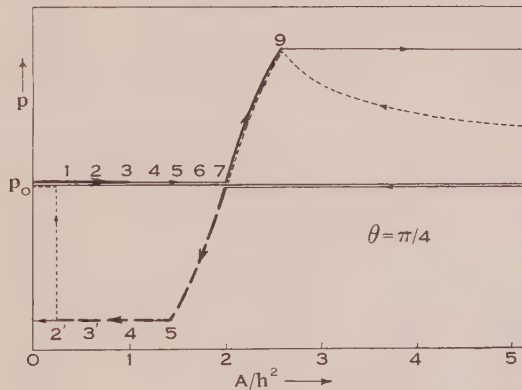


Fig. 6 (b)

Figures 6 (a) and 6 (b) relate to a step profile the same as that of Figure 5 (a) but with a contact angle $\pi/4$. In this case (and when $\theta < \pi/4$) nucleation is easy and the α -phase may be regarded as growing from both A and A' , the diagram remaining symmetrical. The profile remains a straight line until it touches the point P . Simultaneously the two parts touch in the centre and, at constant pressure, growth continues spontaneously until the straight line 7 is obtained. It then continues in equilibrium until stage 9 when the profile leaves P and P' and growth occurs without equilibrium to a large size. The stable (equilibrium) stages of growth are shown by heavy unbroken lines in Figure 6 (b). The evaporation occurs very much as in the previous case, bifurcation occurring when the profile touches AA' . The stable ranges of evaporation are, as before,

shown by heavy broken lines and the equilibrium pressures corresponding to the non-equilibrium stages by fine broken lines. The fine broken lines which might have been given for the growth process have been omitted for clarity. In this case none of the hysteresis is attributable to nucleation at A .

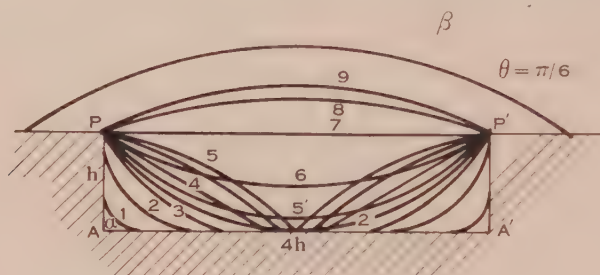


Fig. 7 (a)

Figures 7 (a) and 7 (b) for $0 \leq \theta \leq \pi/6$ should be self-explanatory in the light of the preceding remarks. It will be seen that embryos build up at A and A' as the pressure increases below p_0 . Following bifurcation, the evaporation process is the reverse of the initial stage of the growth process.

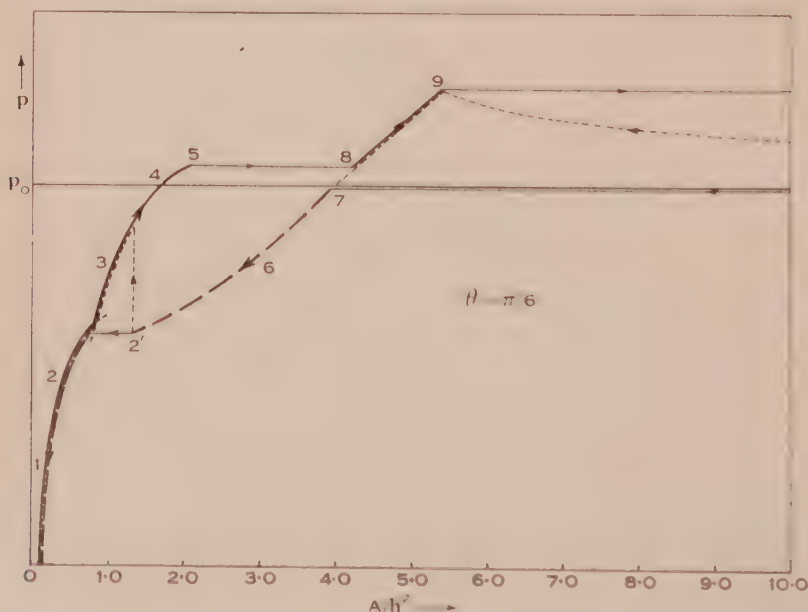


Fig. 7 (b)

It will be clear that situations much more complex can occur with highly imperfect crystal surfaces. However, the basic principle is the same, that nucleation, change of shape, and apparent coalescence give rise to irreversibility. From inspection of Figure 7 (b), for example, it appears that the principal effect results from an apparent nucleation process associated with the separation

of the profile from the points P and P' . However, if the crystal surface is highly imperfect, as is quite likely to be the case when relatively large crystals are subjected to repeated fracture by powdering or when growth pyramids are cut by large numbers of freshly produced screw-dislocation ends, the profile may detach from anchor points such as P and P' only to collide with other obstructions, and having by-passed these may continue to meet others. The number of possibilities is much the larger for the fact that a growing surface phase suffers obstruction not only when it approaches a step on the lower plane but also when it touches the edge on the upper plane. The contact angle in this event will, of course, not be θ but will probably be close to π . The case of Figure 7 (*a*) is only one of an almost infinite number of possibilities in which α -embryos from different sites touch and change shape discontinuously. The sources of colliding embryos may be widely spaced. A particular case not included in the above considerations is that of a rectangular monomolecular recess in an otherwise perfect region of crystal surface. Assuming $\theta > \pi/4$, nucleation of the adsorbate will occur in one corner of this when it will rapidly fill. The removal of this rectangular sheet of α -phase necessitates the nucleation, at a corner, of the β -phase. This process may occur no more easily than the original α -nucleation so that hysteresis for this sheet is greatly enhanced.

Two important questions remain to be discussed. These concern the mobility or immobility of the surface phases and the related rate of transition $\beta \rightarrow \alpha$. Under the conditions of an experiment in which an adsorption isotherm is properly determined, sufficient time is allowed for equilibrium to be established at every relevant pressure, p . That is, the hysteresis does not result from performing the experimental cycle too rapidly. It follows immediately that an active dynamical equilibrium must exist either for the transitions $\text{gas} \rightleftharpoons \alpha$ and $\text{gas} \rightleftharpoons \beta$, or for $\text{gas} \rightleftharpoons \beta$ and $\beta \rightleftharpoons \alpha$, or for all three equilibria. The first possibility does not require the molecules of α - or β -phases to be mobile. However, it is difficult to imagine the molecules of the β -phase immobile as far as surface diffusion is concerned if they are capable, with significant probability, of leaving the surface; this would require an energy barrier to diffusion greater than that for evaporation, assuming that the frequency factors did not differ greatly. The important point is that there is no need to assume that the α -phase is fluid in character in order for the phenomena of apparent coalescence detailed above to occur. When the boundaries of two immobile α -embryos touch, the advance which would be expected on a fluid model occurs by rapid growth from the β - and gas phases. If this growth process is inherently slow then it will be observed that a considerable time is required for the establishment of equilibrium in the hysteresis region. The theory should be equally applicable to physical adsorption either at a perfectly clean crystal substrate or at a surface on which a monolayer of the same or some other gas is already adsorbed chemically. The latter case, perhaps, occurs far more commonly than is generally thought. In practice, it may be that θ varies over a wide range of values for different cases of adsorption. In particular the angle may be widely different for different gases on the same substrate so that the character of the phenomena which occur will

by no means be characteristic of the substrate alone. The contact angle will be the same at two different steps only if they are crystallographically equivalent (σ being isotropic).

When the β -phase is a gaseous film composed of simple molecules, the order of magnitude of the effects to be expected can be estimated. As far as the initial nucleation of the α -phase is concerned, the work of formation of the equilibrium embryo has a value equal to or less than the work of formation of a complete circular equilibrium embryo so that it is sufficient to ask at what pressure a complete circular embryo of α -phase can form and grow with observable probability. A particular case of the nucleation of such a circular plate is the spontaneous formation of a condensed plate of molecules on the perfect face of a crystal of the same substance as substrate. The probability for this process may be calculated from the Becker-Döring theory of formation of crystals in a vapour. As developed by Cabrera and Burton (1949), the theory shows that a vapour pressure roughly twice p_0 is necessary to permit the formation of the new sheet of molecules with observable probability. It may thus be expected that, in some cases, pressures certainly 10 per cent. and possibly as much as 100 per cent. greater than p_0 will be necessary for the initial α -nucleation which occurs at surface defects.

As regards apparent nucleation effects associated with maxima such as that of Figure 3, it is relevant to enquire what minimum radius is associated with a significant excess of the maximum pressure over the value p_0 . The line free energy, σ , may be estimated roughly as follows.

By analogy with Kossel's model of a simple-cubic crystal, the α -embryo may be thought of as a square array of cube-shaped molecules. Those molecules not occupying edge positions are linked to four nearest-neighbours and the four bonds may each be associated with a potential energy U (relative to complete separation on the surface). Taking only potential energy into account, the molecular heat of condensation from the β -phase, λ , is $\lambda = 2U$, and the boundary free energy per molecule is $U/2$. If the linear dimension of a molecule is a then, for present purposes, σ may be taken as $U/2a$, that is $\sigma = \lambda/4a$. Taking the heat of condensation to be 2000 cal. mol. (approximately one fifth the heat of bulk condensation of water vapour) and $a = 3 \text{ \AA}$ (the approximate diameter of a water molecule), a value 1×10^{-6} ergs/cm. is obtained for σ .

From equation (8), with this value of σ and with $T = 300^\circ \text{K.}$, an equilibrium radius of 10^{-6} cm. is found to correspond to a pressure 2 per cent. in excess of p_0 . A radius of 10^{-7} cm. would correspond to an equilibrium pressure 12 per cent. greater than p_0 . However, at this radius thermal fluctuations very probably cannot be neglected as far as the surmounting of the pressure barrier is concerned, especially as the boundary line is only part of a circle so that the change in free-energy associated with a given change in radius is less than the corresponding change for a complete circular α -embryo. As to whether obstacles on the surface may be regarded as separated by a mean distance as small as 10^{-6} cm. or less, little can be said. In the theory of mechanical deformation of metals it has been estimated that about 10^8 dislocation lines terminate on each square centimetre

of surface. If this is correct and a considerable proportion of the dislocation lines are screw axes at the point of contact with the surface, the mean separation of obstructions must at least be considerably smaller than 10^{-4} cm. The grain and sub-grain boundaries associated respectively with grain structure and with mosaic structure also provide obstructions which are essentially of dislocation-type since such boundary surfaces may be resolved into arrays of dislocations. It may thus be concluded that if the mean separation of obstacles is small enough, pressures exceeding p_0 by about 10 per cent. may be necessary to surmount pressure maxima analogous to that of Figure 3. Even allowing for the considerable uncertainty in the estimated values of σ and the mean separation of obstacles, it seems very probable that, apart from initial nucleation, a pressure excess of at least 1 per cent. would be necessary to cover some badly dislocated surfaces with the α -condensate. It is not difficult to see that the evaporation of the α -phase may likewise proceed at an appreciable rate only when the pressure is reduced 1 per cent. or considerably further below p_0 . The two-dimensional phase transition, unlike a phase transition in three dimensions, is obliged to take place on the imperfect surface of a solid and the inter-phase boundary must surmount whatever obstacles the surface provides, obstacles that will not yield to the forces exerted by the advancing boundary.

IV. CONCLUSION

The arguments put forward show that phase transitions in two dimensions provide a possible explanation for the hysteresis of adsorption of gases at some types of solid surface. Indeed, the effects described almost certainly do occur; the real question is one of the relative magnitudes of these and other possible mechanisms. Three-dimensional capillary condensation is known to occur in some cases (Carman and Raal 1951), but many unexplained examples of hysteresis exist. In the present theory the hysteresis is associated with irreversible effects in the growth of a new surface phase. These may be particularly associated with the early stages of its formation. Since the whole of the α -phase must originate from embryonic growth processes its formation may occur in just such a manner as to give rise to hysteresis.

The calculation of expected adsorption isotherms from the above considerations as basis would involve having some idea of the nature and quantity of the surface defects, and the theory, to be realistic, would have to be statistical in character. Insufficient is known about the state of surfaces for which data have been reported to warrant the attempt. In any case calorimetric data are also required and the energy of the boundary line must be estimated. However, the following predictions of the theory are in agreement with experiment:

1. The hysteresis is associated with a phase transition (for example, Emmett and De Witt 1943; Gregg and Maggs 1948).
2. Surfaces which are not porous can exhibit hysteresis (Bangham and Mosallam 1938; Pierce and Smith 1950).
3. The time required for equilibrium to be established in the hysteresis region can be relatively very large (Gregg 1949).

4. If a surface is highly imperfect the observed transition will appear diffuse (the theory suggests this particularly when θ is small). The diffuse transition of first order is very common.
5. The observed sensitivity of hysteresis to contamination may be interpreted as a manifestation of the constraints which the contamination offers to the advancing boundary lines and its effect on the contact angles at existing obstacles.

V. REFERENCES

- BANGHAM, D. H., and MOSALLAM, S. (1938).—*Proc. Roy. Soc. A* **165** : 552; **166** : 558.
BURTON, W. K., CABRERA, N., and FRANK, F. C. (1951).—*Philos. Trans.* **243** : 299.
CABRERA, N., and BURTON, W. K. (1949).—*Discuss. Faraday Soc.* No. 5 : **40**.
CARMAN, D. C., and RAAL, F. A. (1951).—*Nature* **167** : 112.
EINSTEIN, A. (1910).—*Ann. Phys. Lpz.* **33** : 1275.
EMMETT, P. H., and DE WITT, T. W. (1943).—*J. Amer. Chem. Soc.* **65** : 1253.
FRANK, F. C. (1949).—*Discuss. Faraday Soc.* No. 5 : 48.
FRENKEL, J. (1924).—*Z. Phys.* **26** : 117.
GREGG, S. J. (1949).—"Surface Chemistry." p. 205. (Butterworth : London.)
GREGG, S. J., and MAGGS, F. A. P. (1948).—*Trans. Faraday Soc.* **44** : 123.
KATZ, S. M. (1949).—*J. Phys. Coll. Chem.* **53** : 1166.
PIERCE, C., and SMITH, R. N. (1950).—*J. Phys. Coll. Chem.* **54** : 784.
TURNBULL, D. (1950).—*J. Chem. Phys.* **18** : 198.
VOLMER, M. (1929).—*Z. Elektrochem.* **35** : 555.
VOLMER, M. (1939).—"Kinetik der Phasenbildung." (Steinkopf : Leipzig.)
ZSIGMONDY, R. (1911).—*Z. anorg. Chem.* **71** : 356.

A STUDY OF THE ANNUAL SOIL TEMPERATURE WAVE

By E. S. WEST*

[*Manuscript received November 26, 1951*]

Summary

Soil temperatures recorded at Griffith over an 8 year period at a depth ranging from 1 in. to 8 ft. have been examined and compared with air temperatures.

The observed fluctuations in the soil temperatures fit closely the theoretical equation for the propagation of a simple harmonic temperature wave into the soil.

The diffusivity of the soil has been deduced and compared with values found by other workers in other localities.

The annual wave of the daily mean temperature at the surface of the soil has been deduced and compared with the annual wave of the daily mean air temperature and the differences in the means, amplitudes, and phase displacements have been discussed.

I. INTRODUCTION

The temperature of the soil and its variations are of interest to those studying plants and animals growing in or on the soil and to those interested in the weather and climate. Records of soil temperature down to a depth of 8 ft. have been kept at the Irrigation Research Station, Griffith, and in this paper the summarized data are presented and discussed. Attention is directed mainly to the annual temperature wave.

II. SITUATION, CLIMATE, AND SOIL

The following data on the location, climate, and soil are given as they are relevant to the soil temperature data.

Griffith is situated at latitude $34^{\circ} 17' S.$ and longitude $146^{\circ} 03' E.$ It is 300 miles inland from the sea on an extensive plain. The climate is between a mediterranean and semi-continental type. It is just outside the wheat belt for dry-land farming, but oranges, grapes, deciduous fruit, and rice grow well under irrigation. The average annual rainfall is 15 in.

The soil belongs to the Grey and Brown soils of the major soil groups and is classified as a Hanwood Sandy Loam (Taylor and Hooper 1938). The A horizon 9 in. deep is a sandy loam. The B1 horizon from 9-17 in. is clay. Below this is clay with limestone concretions.

* C.S.I.R.O. Irrigation Research Station, Griffith, N.S.W.

III. DESCRIPTION OF RECORDING INSTRUMENTS

Thermograph records were taken at depths of 1 and 6 in., 1 and 2 ft., and records from mercury-in-glass thermometers at 4, 6, and 8 ft.

The recording instruments were placed in position as follows.

A rectangular plot of land 5 by 4 yd. was kept untilled but free of weeds and traffic. The recording "points" of the instruments were placed at the appropriate depths in the centre of this plot. The ground for some yards around this plot was also kept free of weeds. The records, therefore, refer to soil temperatures under bare, uncultivated land.

A concrete-lined pit was made on the south side of the plot. The pit was 6 ft. long, 3 ft. wide, and 10 ft. deep, the concrete walls being 4 in. thick. The 3 ft. side of the pit was centred on the south side of the plot. The top of the pit was closed with trap doors well insulated with cork dust.

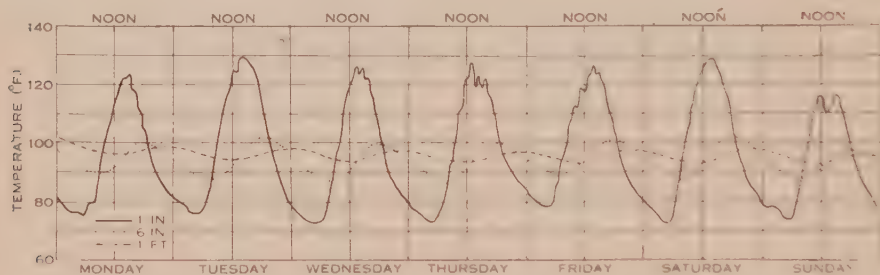


Fig. 1.—Soil temperatures, January 16-22, 1939.

This pit housed the recording instruments. The thermographs were Negretti and Zambra instruments. Their bulbs were placed at the appropriate depth 6 ft. to the north of the pit, that is, in the centre of the plot. The leads came in horizontally from the bulb to the pit.

The temperatures at the 1, 6, and 8 ft. depths were recorded by means of mercury-in-glass thermometers with stems 2 metres long. The stems of these thermometers were encased in wood and the bulbs were protected by a pointed metal sheath. To instal these thermometers horizontal holes were bored from the pit of just sufficient bore to take the wooden encasement of the stems. After pushing the thermometers to the full length of the bores the instruments were lightly hammered to force the pointed steel sheaths into the soil at the end of the bore. These mercury-in-glass thermometers were very accurate. They were graduated in fifths of a degree Fahrenheit.

The objection may be raised that bare soil is a rather unnatural condition. However, it is less unnatural than tilled soil, and the natural herbage is so varied in place and time in this locality that it would be difficult to decide on what is a "typical" cover, and it would be still more difficult to maintain it. "Bare" soil at least has the virtue of being readily described and easily reproduced.

The presence of the pit would affect the soil isotherms in its vicinity. However, the results shown in Figure 4 suggest that the interference was negligible at the positions of the thermometer bulbs.

IV. RECORDS AVAILABLE

Reasonably complete records are available for the 8 calendar years 1934-41, and the data for this period are used in this discussion.

Figure 1 shows typical January (summer) soil thermographs for the 1, 6, and 12 in. depths, and Figure 2 shows typical July (winter) thermographs. The general characteristics of these graphs are well known and will not be discussed here except in so far as they are relevant to the discussion on the annual temperature wave.

(a) *Compilation of Records*

For the purpose of the harmonic analysis of the annual temperature waves weekly means were compiled for the 52 weekly periods from January 1 to

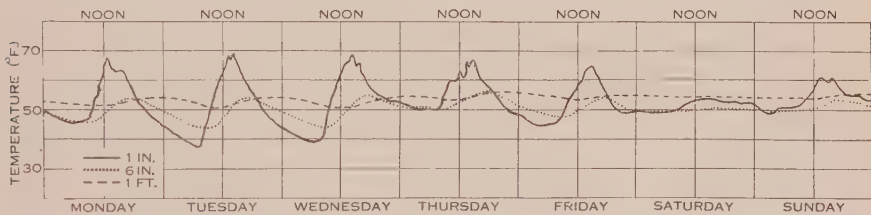


Fig. 2.—Soil temperatures, July 20-26, 1936.

December 30. This left December 31 as an odd day. In obtaining the mean daily temperatures the mean temperatures for this day were halved and one half allotted to the sums of the preceding and one half to the sums of the succeeding week. In a similar way the temperatures for February 29, 1936 and 1940 were included in the means for that period.

The weekly period was chosen as it was sufficiently fine to permit reasonable accuracy in fitting the harmonics, while at the same time not making the calculation too cumbersome. For convenience the period from 9 p.m. to 9 p.m. was used for obtaining the daily means. Thus the year was measured from 9 p.m. December 31.

The mean daily temperatures were obtained as follows.

(i) *One In. Depth.*—The shape of the diurnal curve is such that halving the sum of the maximum and minimum does not give a good estimate of the mean. A preliminary examination showed that the mean of the four readings 6 a.m., noon, 6 p.m., and midnight gave a close approximation to the true mean found

by summing under the continuous curve. Using readings at 4-hourly or 2-hourly intervals did not appreciably improve the estimate. These four readings were therefore used. The figures for the weekly means for this depth are, therefore, derived from 224 observations (4 readings for 7 days of 8 years).

The recording instruments were checked before installing and were found to record accurately over the whole range recorded. However, no provision was made to check the instrument after installation. A further difficulty arose from the fact that the recording pen arm for the 1 in. depth had to be lowered 30 °F. during summer to keep it on the scale. The precaution was taken of marking the position of the pen before and after these adjustments so that the amount of adjustment was recorded. Nevertheless, the possibility exists that an error in setting may have accumulated over the years. As the amplitude of the diurnal wave at this depth is so large, the observations as read from the charts are subject to rather large errors compared with those for the other depths, but owing to the large number of observations that go into one weekly mean as recorded, the mean measured by the method adopted would be quite good. However, there are likely to be large systematic errors (relative to the readings for the other depths). These would affect the annual means but probably not the amplitudes or phases of the periodic curves.

(ii) *Six In. Depth.*—The daily thermograph for the 6 in. depth is smoother than that for the 1 in. depth and is markedly damped but still departs markedly from the sine curve. However, half the sum of the minimum and maximum gives a close approximation to the mean and this was used.

(iii) *One Ft. Depth.*—The daily thermograph at this depth closely approximates a sine curve and half the sum of the minimum and maximum was used to determine the mean.

(iv) *Two Ft. Depth.*—The daily wave at this depth is almost damped out. The reading at 9 a.m. was taken for this depth.

(v) *Four, Six, and Eight Ft. Depths.*—These thermometers were read at 9 a.m. each morning. Owing to a rise in the water table the 8 ft. thermometer was flooded for a period of about 1 year, and records are not available for this period. Rather than restrict the whole analysis to a 7 year period the 8 ft. temperatures were estimated for this period. For each week the mean temperature for the 8 ft. depth was plotted against that for the 6 ft. depth for the 7 years that the complete data are available. The plots fell close to straight lines and the missing 8 ft. depth figure was read off using the figure for the 6 ft. depth for that year. The error involved in this estimation would be quite small.

(vi) *Air Temperature.*—Temperatures were taken in a standard Stevenson screen situated close to where the soil temperatures were taken. Maximum and minimum and 9 a.m. readings were available. A preliminary examination

based on 48 sample days in January and 48 in July gave the following comparisons of estimates of the mean compared with the mean as calculated by summing under the curve.

Readings	January		July	
	Mean Difference (°F.)	Standard Deviation	Mean Difference (°F.)	Standard Deviation
(max. + min.)/2 ..	+0.36	1.961	-0.19	0.889
9 a.m. ..	+0.02	4.021	+3.53	5.77

Half the sum of the minimum and maximum readings gives a more precise estimate of the means than the 9 a.m. reading and this estimate was therefore used.

V. ANALYSIS OF DATA

Harmonic analyses were made of the summarized temperature data for each depth of the soil and for the air. The following are the equations for the fundamental and first harmonics for the soil temperature at the 1 in. depth and the air temperature.

$$\theta_{1 \text{ in.}} = 66.44 + 21.43 \sin \left(\frac{2\pi t}{T} + 1.377 \right) + 1.514 \sin \left(\frac{4\pi t}{T} + 0.609 \right),$$

$$[\theta\theta] = 11,990.557 \quad [v_1v_1] = 209.291 \quad [v_2v_2] = 149.661,$$

$$\theta_a = 61.38 + 14.075 \sin \left(\frac{2\pi t}{T} + 1.323 \right) + 1.21 \sin \left(\frac{4\pi t}{T} + 0.108 \right),$$

$$[\theta\theta] = 5,285.777 \quad [v_1v_1] = 143.828 \quad [v_2v_2] = 105.734,$$

where $\theta_{1 \text{ in.}}$ and θ_a = the mean daily soil temperature at 1 in. depth and the mean daily air temperature respectively,

t = the time in days measured from 9 p.m. December 31,

T = the period, that is, taken as 364 days.

The angles are given in radians, and

$[\theta\theta]$ = the sum of squares of the departure of the "observed" temperatures from the mean. By "observed" temperatures is meant the 8 years' means of the weekly periods,

$[v_1v_1]$ = the sum of the squares of the residuals after fitting the fundamental term,

$[v_2v_2]$ = the sum of squares of the residuals after fitting the first harmonic term.

It is seen that the observed data for both the soil and air temperatures fit closely a simple sine curve. The coefficients of the first harmonic terms are quite small. Most of the total variance is accounted for by the fundamental term the first harmonic not reducing the residual variance by an appreciable

amount. As one would expect, simple sine curves fit the temperature for the lower depths still more closely so that very little is gained in considering the higher harmonics and the following discussion will be confined to the simple sine curves.

The following are the equations fitted for each depth :

$$\left. \begin{aligned} \theta_{1\text{ in.}} &= 66.44 + 21.43 \sin \left(\frac{2\pi t}{T} + 1.377 \right) & [vv] &= 209.291 \\ \theta_{6\text{ in.}} &= 68.12 + 20.22 \sin \left(\frac{2\pi t}{T} + 1.345 \right) & [vv] &= 101.841 \\ \theta_{1\text{ ft.}} &= 70.93 + 18.71 \sin \left(\frac{2\pi t}{T} + 1.261 \right) & [vv] &= 75.743 \\ \theta_{2\text{ ft.}} &= 69.35 + 15.27 \sin \left(\frac{2\pi t}{T} + 1.125 \right) & [vv] &= 28.684 \\ \theta_{4\text{ ft.}} &= 69.68 + 11.17 \sin \left(\frac{2\pi t}{T} + 0.827 \right) & [vv] &= 8.645 \\ \theta_{6\text{ ft.}} &= 69.66 + 8.43 \sin \left(\frac{2\pi t}{T} + 0.565 \right) & [vv] &= 4.515 \\ \theta_{8\text{ ft.}} &= 69.54 + 6.50 \sin \left(\frac{2\pi t}{T} + 0.347 \right) & [vv] &= 2.006 \\ \theta_a &= 61.38 + 14.07 \sin \left(\frac{2\pi t}{T} + 1.323 \right) & [vv] &= 143.828 \end{aligned} \right\} \dots (1)$$

The sum of the residuals $[vv]$ are given after each equation.

As explained, we can expect the mean temperatures for the depth 4, 6, and 8 ft. to be determined with a high degree of accuracy. The means of the temperatures at 1 and 6 in., 1 and 2 ft. being based on thermograph records would not be so precise, but the 2 ft. mean would be quite good, as the only error would be the instrument error: that for 1 ft. would be fairly good as the daily wave is small at this depth and the error of estimating the daily mean would be small: that for 6 in. would be less accurate, as the daily wave, although small, is not a sine curve; while that for 1 in. would be the least accurate owing to the difficulties of estimating the mean from the irregular curve of a very large amplitude, and owing to the further source of error that arose from adjusting the pen arm of the instrument every spring and autumn.

Considering the expected accuracy of the determination of the mean temperature at different depths, one would not be justified in making inferences from the differences in the mean temperatures as shown above, and it is reasonable to conclude that the mean temperature at the various depths was approximately 69.6°F .

Figure 3 shows the curves fitted for the 1 in. and 8 ft. annual temperature waves together with the observed weekly means. It will be noticed that at the 1 in. depth the temperature is much more variable in the summer than in the winter. It will also be noticed that the means for the 8 year period show a subsidiary wave with a period of about 6 weeks. Owing to short-period weather changes the temperature of the surface of the soil fluctuates about the annual

temperature trend leading to waves of temperature with periods varying from about 5 to 10 days. Averaging this effect over a period of 8 years reduces the amplitudes but lengthens the periods of these waves, but 8 years is probably not a sufficiently long period to remove them entirely. Though it might be expected that a sufficiently long period would cause these fluctuations to disappear, the possibility must not be overlooked that they may, in fact, be genuine cycles superimposed on the annual trend, similar in nature to the well-known Buchan cycle.

It will be seen that for the 8 ft. depth the observed points fall almost exactly on the sine curve. The departures from the smooth curve seen in the plot for the 1 in. depth decrease more or less logarithmically with depth as one might expect. If the sole source of these departures was due to the fluctuations in

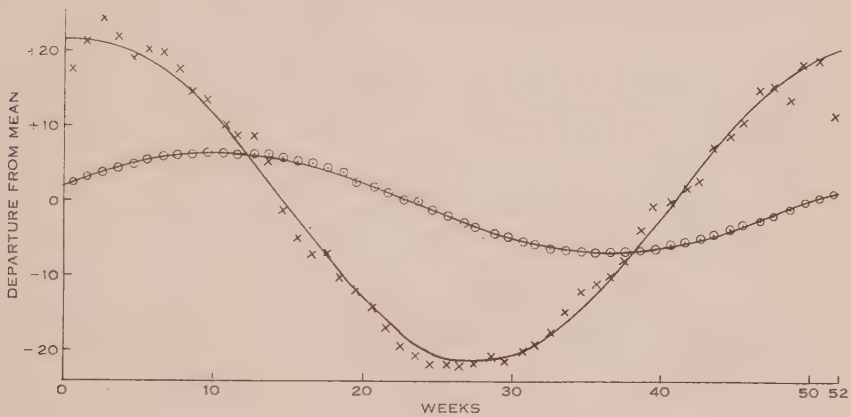


Fig. 3.—Curves fitted for the 1 in. (x) and 8 ft. (o) annual temperature wave.

the weather one would expect $[vv]$ to be an inverse logarithmic function of the depth. This relation holds fairly well for the lower depths but the curve $\log [vv] = b - cx$ curves up for the top three depths. This is what was expected as the mean temperatures for the 1 ft. and 6 in. depths were less accurately estimated than those for the lower depths and the means for the 1 in. depth were subject to still greater errors of estimate.

The coefficient of the sine term gives the amplitudes of the curves, and the constant in the sine term gives the phase. As explained, one can expect these to be more accurately determined than the means for the surface depths as the instrument errors and errors of estimating the means would probably not affect these statistics.

These equations can now be considered in terms of the theoretical equation for the propagation of a simple harmonic temperature wave in a uniform medium.

If the soil has a uniform heat diffusivity, and a varying temperature in the form of a simple harmonic heat wave is applied to the surface, the propagation of the heat wave downwards is given by

$$\theta = \theta_0 e^{-2\pi x/\lambda} \sin 2\pi \left(\frac{t - \gamma}{T} - \frac{x}{\lambda} \right), \quad \dots \dots \dots (2)$$

where θ = the temperature in $^{\circ}\text{F.}$ measured from the mean,

θ_0 = the amplitude at the surface,

x = depth in ft.,

λ = wavelength, that is, the distance between two depths in the soil at which the maximum or minimum occurs simultaneously,

t = the time in days (measured from 9 p.m. December 31),

T = the period (taken as 364 days),

γ = the phase displacement in days at the surface.

The amplitude at any depth is given by the exponential term and the phase at any depth is given by the second term in the brackets of the sine term.

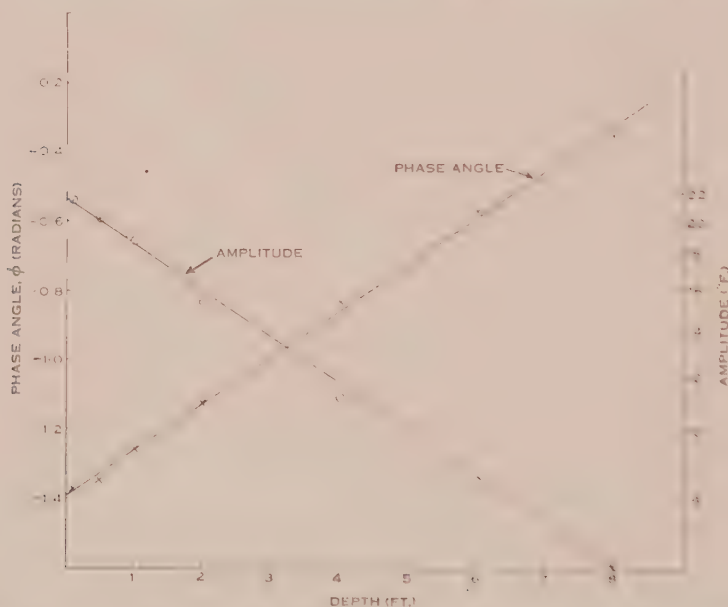


Fig. 4.—Relation between phase angle and depth, and amplitude and depth.

The diffusivity is given by

$$k = \frac{K}{c} \frac{\lambda^2}{4\pi T^2} \quad \dots\dots\dots (3)$$

where K is the heat conductivity and c is the thermal capacity per unit volume.

This gives a method for determining k .

From the sine term in (2) we have

$$\varphi = \frac{2\pi\gamma}{T} + \frac{2\pi x}{\lambda}$$

where φ = the phase displacement angle, that is, the constants in the sine terms of (1).

Figure 4 shows the values of φ in (1) plotted against x . The regression equation is

$$\varphi = -1.39213 + 0.13422x \quad \dots\dots\dots (4)$$

from which is obtained the following estimates :

$$\gamma = -80.65 \text{ days}, \quad \lambda = 46.81 \text{ ft.}$$

From the exponential term in (2) we have

$$A = \theta_0 e^{-2\pi x/\lambda},$$

where A is the coefficient of (1) that is the amplitude, or

$$\log_e A = \log_e \theta_0 - \frac{2\pi}{\lambda} x.$$

Figure 4 shows the values of $\log_{10} A$ in (1) plotted against x . The regression equation is

$$\log_{10} A = 1.3314 - 0.06671x \quad \dots\dots\dots (5)$$

from which is obtained the following estimates :

$$\theta_0 = 21.45^\circ \text{F.}, \quad \lambda = 40.90 \text{ ft.}$$

The agreement between the two estimates of λ is reasonably good.

The diffusivity will vary with depth according to the physical properties of the soil. It will also vary with depth and time according to the moisture content of the soil. The effect of moisture is to increase the thermal capacity and also the conductivity. The net effect is for the diffusivity at first to rise with increasing moisture content, pass through a maximum and fall (Patten 1909). These results, therefore, refer to average conditions of moisture content over the period studied.

The relations between φ and x and between $\log_{10} A$ and x shown in Figure 4 fit straight lines quite satisfactorily. We can, therefore, assume that the average diffusivity is reasonably constant over the depth studied.

Numerical values were calculated for the constants θ_0 , γ , and λ in equation (2) to give the equation of best fit and the following values were found :

$$\theta_0 = 21.34, \quad \gamma = -81.92, \quad \lambda = 41.96. \quad \dots\dots\dots (6)$$

The generalized equation (2) then becomes

$$\theta_0 = 21.34 e^{-2\pi x/41.96} \sin 2\pi \left(\frac{t+81.92}{T} - \frac{x}{41.96} \right). \quad \dots\dots (7)$$

The sum of squares of the deviations of the observed values from the mean can be apportioned as follows :

ss. residuals from the particular curves for each depth (equation (1))	430.726
ss. of deviations of generalized equation from the particular equations	60.508
Total ss. residuals from generalized equation	491.234
ss. of generalized equation	43,768.515
Total ss.	44,199.241

The use of the generalized equation only slightly increases the sum of the squares of the residuals, showing that the generalized equation fits the observed data closely.

(a) *Evaluation of Diffusivity*

Using the value of λ given in (6) in equation (3) we have the diffusivity $k = 0.3850$ ft.²/day or in c.g.s. units 0.004319 cm.²/sec. Table 1 shows how this compares with values found by other workers.

TABLE 1
DIFFUSIVITY VALUES COMPARED WITH THOSE OF OTHER WORKERS

Author	Soil Type	Locality	Measurement	k (cm. ² /sec.)
Present work ..	Clay	Griffith	Annual wave	0.00432
Rambaut (1916) ..	Gravel	Oxford	Annual wave	0.00758
Kelvin*	Trap rock	Calton Hill	Annual wave	0.00786
	Sand	Calton Hill	Annual wave	0.00871
	Sandstone	Craigleith	Annual wave	0.02310
		Quarry		
Everett†	Gravel	Greenwick	Annual wave	0.01249
Wright (1931) ..	Light loam	Kew	Annual and diurnal waves	0.0044
Wright (1931) ..	Light loam	Kew	Annual and diurnal waves	0.0066
Wright (1931) ..	Sand	Potsdam	Annual wave	0.0109
Patten (1909) ..	Sands	Washington, D.C.	Laboratory measurement	0.0026-0.008
Patten (1909) ..	Loams	Washington, D.C.	Laboratory measurement	0.003-0.007
Patten (1909) ..	Clay	Washington, D.C.	Laboratory measurement	0.0025-0.0045

* Reported by Preston (1919).

† Reported by Rambaut (1916).

Patten's results were for a variety of soils at a range of moisture content. The lower limits of his values were for dry soils and the highest limits were for soils at the moisture content of maximum diffusivity. The values of the diffusivity at Griffith are similar to those reported for loams and clays by other workers. Sands and rocks tend to have a higher diffusivity.

(b) *Relation of Air Temperature to Surface Soil Temperatures*

Putting $x = 0$ in equation (7) and introducing the mean temperature as a constant we have the temperature at the surface in degrees Fahrenheit as follows :

$$\theta_s = 69.6 + 21.34 \sin 2\pi \left(\frac{t + 81.92}{T} \right). \dots\dots\dots (8)$$

For comparison the equation for the air temperature expressed in the same form is

$$\theta_a = 61.4 + 14.08 \sin 2\pi \left(\frac{t + 76.64}{T} \right). \dots\dots\dots (9)$$

The surface of the soil receives short-wave radiation from the Sun and some long-wave radiation from the atmosphere. It loses heat by long-wave radiation into space and by conduction to the air. There is also a downward diffusion of heat into the soil in summer and autumn and an upward diffusion of heat to the surface during winter and spring. The conduction of heat to the air is not true conduction but is due mainly to the mass movement of warm air upwards due to turbulence caused by wind. The surface of the soil is thus the centre from which heat is received and lost, and controls the temperature of the air above and soil beneath. Conduction is the only important process in the distribution of heat within the soil. The transfer of heat by movement of water and air within the soil is of minor importance. The transfer of heat to and from the air is more complex. Water evaporating from foliage would cool the air; but it is probable that the difference between the average air temperature and the average temperature of the surface of the soil is due to the moving in of cold air from cooler regions. This moving in of cold air may be from a distant or a local source. For example, if the wind blows more frequently from a cooler region say from the south, than from a warmer region say from the north-west, one would expect the air to have a lower mean temperature than the soil. Again the observation plot which was on bare soil may be expected to have a higher mean surface temperature than adjacent grass- or tree-covered surface, as found by Johnson and Davies (1927), which latter would then be a source of cold air.

Comparisons between the mean air temperature and the mean soil temperature for different localities are shown in Table 2. At the first three localities

TABLE 2
COMPARISON OF AIR AND SOIL TEMPERATURES AT DIFFERENT LOCALITIES

Locality	Mean Soil Temperature (°F.)	Mean Air Temperature (°F.)
Auburn, Alabama (Fitton and Brooks 1931) ..	64.9	63.6
Oxford, England (Rambaut 1916)	50.7	49.2
Urbana, Illinois (Fitton and Brooks 1931)	53.0	51.0
East Lansing, Michigan (Fitton and Brooks 1931) ..	50.6	47.3
Anaheim, California (Bliss 1944)	64.6	58.7
Griffith, N.S.W.	69.6	61.4
Indio, California (Bliss, Moore, and Bream 1942) ...	67.4	71.9

the soil and air temperature are approximately the same. The small differences could possibly be explained by the methods adopted in measuring the temperatures. For the next three localities the mean air temperature is appreciably below the mean soil temperature. The cool air temperature at Anaheim is

ascribable to the cool sea breezes of the locality. At the last locality, Indio, the mean air temperature is higher than the mean soil temperature and this is explained by winds blowing from hot interior deserts (Bliss, Moore, and Bream 1942).

(c) Phase Displacement

The longest day occurs on December 22. From equation (8) it is deduced that the curve of the temperature of the surface of the soil is at its mean position when $t = -81.92$, that is, 82 days before January 1. The maximum temperature occurs $90 \times 364/360 = 91$ days after this, that is, on January 9. In a similar way it is deduced from equation (9) that the maximum air temperature occurs on January 14. That is, that there is a lag of about 18 days in the maximum of the soil surface temperature curve behind midsummer and a further lag of 5 days in the maximum air temperature curve.

The amplitude of the air temperature wave is about two-thirds that of the soil surface temperature wave, but owing to the difference in the mean temperatures the winter temperatures are about the same, the summer air temperatures being about 14°F. lower than the summer soil surface temperatures.

We are, of course, discussing the trends in average daily temperatures. The daily maximum temperatures in the summer are much higher.

VI. ACKNOWLEDGMENT

The author wishes to acknowledge the help of Mr. W. C. Swinbank, Section of Meteorological Physics, C.S.I.R.O., for reading the typescript and making helpful suggestions.

VII. REFERENCES

- BLISS, D. E. (1944).—Air and soil temperatures in a California citrus orchard. *Soil Sci.* **58**: 259-74.
- BLISS, D. E., MOORE, D. C., and BREAM, C. E. (1942).—Air and soil temperatures in a California date garden. *Soil Sci.* **53**: 55-64.
- FITTON, E. M., and BROOKS, C. F. (1931).—Soil temperatures in the United States. *Mon. Weath. Rev.* **59**: 6-16.
- JOHNSON, N. K., and DAVIES, E. L. (1927).—Some measurements of temperatures near the surface in various kinds of soils. *Quart. J. R. Met. Soc.* **53**: 45-59.
- PATTEN, H. E. (1909).—Heat transference in soils. Bull. U.S. Div. Soils No. 59.
- PRESTON, T. (1919).—"The Theory of Heat." (MacMillan & Co.: London.)
- RAMBAUT, A. A. (1916).—Underground temperatures as determined by platinum resistance thermometers. Ratcliffe Observatory, Oxford. *Astr. Met. Obs.* **51**: 103-123.
- TAYLOR, J. K., and HOOPER, P. D. (1938).—A soil survey of the horticultural soils in the Murrumbidgee Irrigation Areas, N.S.W. Coun. Sci. Industr. Res. Aust. Bull. No. 118.
- WRIGHT, H. L. (1931).—The variation of soil temperature below turf: A discussion of observations at Kew Observatory. *Mem. R. Met. Soc.* **4**: 1-18.

THE RATE OF NATURAL EVAPORATION FROM SHALLOW PONDS

By J. FERGUSON*

[Manuscript received October 1, 1951]

Summary

The meteorological factors determining the rate of natural evaporation from "free" water surfaces are: the radiant energy supply from the Sun and atmosphere, the air temperature in contact with the surface, the absolute humidity of the air, and the wind velocity and turbulence over the surface. The relevant relationship of these factors has been expressed in a differential equation, whose solution, given proper numerical instantaneous values of the determining factors and their variation with time, yields the value of the depth of liquid evaporated in a given period from small areas of shallow ponds of different depths. Using different values of the parameters, a number of solutions of the equation were obtained by means of a differential analyser. It is shown that, provided the depth of the water is greater than about 16 cm., a simpler equation using mean values of the radiant energy flux and the air temperature over short periods of time yields almost identical results.

The use of this equation for the prediction of evaporation from shallow ponds exposed to known climatic conditions is discussed.

I. INTRODUCTION

The investigation reported here is a study—carried out in 1945—of the factors governing the rate of evaporation of water from shallow solar ponds. It was undertaken with the primary object of devising a method of calculating from the necessary and given meteorological data the amount of water evaporated in a specified length of time from a shallow pond of known dimensions exposed to the influences of Sun, sky, and wind.

The subject of evaporation from "free" water surfaces has, of course, received much attention in the past from physicists, civil engineers, agriculturists, oceanographers, etc. Most of the investigations performed have been concerned essentially with the determination of the value of the coefficient, k , in the Dalton equation

$$\frac{dw}{dt} = Ak(p_w - p_a)$$

at different wind velocities and different degrees of atmospheric turbulence.

In this equation dw/dt is the mass rate of evaporation from a free water surface of area, A , into air containing water vapour at a pressure p_a , and p_w is the vapour pressure of the evaporating water.

It is, however, clear that this equation alone cannot be used to calculate an evaporation rate without a knowledge of p_w , that is, without a knowledge of

* Research Department, Imperial Chemical Industries Ltd. (Alkali Division), Northwich, England.

the water surface temperature and how this changes as evaporation proceeds—data which, of course, are not known *a priori*. In natural evaporation the water temperature and its variation with time are functions of the meteorological conditions governing the rate of gain and loss of energy from the evaporating water. Energy enters as short- and long-wave radiation from the Sun, sky, and atmosphere, and, sometimes, as sensible heat from hotter air in contact with the water surface. Energy is lost as long-wave radiation, sometimes as sensible heat conveyed to colder air above the pond, and finally as latent heat carried off by the water vapour evaporated. The problem of calculating “free” evaporation rates cannot be solved, at least in its most general form, without constructing a heat balance.

The heat balance type of approach to the evaporation problem has received much less attention than the aerodynamic one. Its importance was particularly stressed by Cummings (1921). Later, Cummings and Richardson (1930) put forward an expression, based on a heat balance, for the calculation of the amount of water evaporated in a given time from an open water surface. Unfortunately their formula contains Bowen's (1925) coefficient, which requires for its evaluation a knowledge of the water temperature.

Recently, Penman (1948) has published a treatment of the evaporation problem based on the heat balance, in which he derives an expression for evaporation rate which does not involve the water temperature. This is rendered possible by using an approximation, viz. by postulating that, within the ranges of air-water temperature differences likely to be encountered in practice, the vapour pressure-temperature curve of water may, with sufficient accuracy, be regarded as a straight line, which is the tangent to the vapour pressure curve at the air temperature.

Penman, like Cummings and Richardson, in deducing his equation, assumes constant values of the heat flows to and from the evaporating surface during the relevant time period. The diurnal variation of the intensity of the incoming radiation and of the air temperature is neglected, and, in the practical application of the formula, mean values of the determining meteorological factors are used to calculate the expected evaporation over periods ranging from a day to a month. In the present paper, the instantaneous values of variable factors determining the heat balance of an evaporating pond are combined in a differential equation, which is then integrated. This procedure has the advantage of a greater degree of theoretical exactitude and throws into clear light the various assumptions which are required to render the problem amenable to treatment.

II. THE HEAT BALANCE IN NATURAL EVAPORATION

Before deriving the evaporation equation, it is desirable first to list the various flows of energy to and from an evaporation pond, and to give some idea of their possible orders of magnitude.

Sensible heat communicated by contact with the overlying air, and radiant energy from the Sun, sky, and atmosphere, are the two and only sources of energy available for evaporating water from “free” natural surfaces. Sensible

heat flows from the air to the pond only, of course, so long as the air temperature exceeds the water temperature. It is known that often in certain climates, at certain times of the day, pond water is hotter than the air over it.* At such times the sensible heat flow is a loss. The heat loss terms consist of this sensible heat loss, the heat consumed in evaporation, and the loss due to radiation of long-wave energy by the pond. The earth underneath the pond may be regarded as an insulator.

(a) The Radiant Energy Supply

The radiant energy supply consists of direct short-wave radiation from the Sun, scattered short-wave radiation from the sky and clouds, and long-wave radiation from the atmosphere. The rate of supply of the total radiant energy to a unit surface area on the Earth's surface is, of course, measurable by suitable instruments. Unfortunately, detailed records of this flux are not often available. From cloudless skies the maximum intensity of short-wave radiation received at noon, in low latitudes, may amount to more than 90 cal./cm.²/hr. As an example of the variation of the horizontal intensity with the Sun's zenith distance, the following results of Kimball (1928) based on measurements under clear sky conditions may be quoted :

Solar zenith distance	0°	60°	75°·7	80°·7	84°·7	88°·2
Air mass	1·0	2·0	4·0	6·0	10·0	21·0
Horizontal intensity (cal./cm. ² /hr.)	86·4	37·2	15·8	8·6	3·5	0·6

The long-wave radiation emitted from the air is also measurable in isolation by suitable instruments. It is due almost entirely to the water vapour content of the air (Brunt 1944). At atmospheric temperatures water vapour is not a black-body radiator. According to Brunt the downcoming radiation from the air can be calculated from a knowledge of the air temperature and absolute humidity by means of the equation

$$R/\sigma T^4 = 0.52 + 0.065\sqrt{e}, \dots\dots\dots (1)$$

where R is the downcoming radiation in cal./cm.²/sec., T the absolute temperature of the air, σ Boltzmann's constant, and e the absolute humidity of the air in millibars.

Using this expression the following values of the downcoming radiation incident upon a horizontal surface have been calculated for a few air temperatures and absolute humidities, p_a , expressed in mm. of mercury :

Air temperature (°C.)	5	10	15	20	25
Radiation (cal./cm. ² /hr.) ($p_a = 5$ mm.)	20·4	21·9	23·5	25·1	26·9		
Radiation (cal./cm. ² /hr.) ($p_a = 15$ mm.)	24·0	25·8	27·7	29·6	31·7		

Where no instrumental records are available an estimate may be made of the total radiation (short- and long-wave), received per day at any locality on the Earth's surface, from a knowledge of the hours of sunshine, cloudiness, air

* Symposium : *Trans. Amer. Soc. Civ. Engrs.* **99** : 671 (1934).

temperature, and humidity. The necessary formulae are listed by Penman (*loc. cit.*). Such calculated figures do not, of course, give the instantaneous values of the radiation flux. The short-wave radiation received may vary according to cloudiness etc. from zero to 900 cal. cm.² 24 hr. As an example of the long-wave radiation flux, the flux under cloudless sky conditions from air at 20 °C. of absolute humidity 10 mm. is about 660 cal. cm.² 24 hr. The maximum figure for the total daily radiation likely to be received at a suitable location on the Earth's surface is therefore of the order of 1550 cal. cm.².

Not all the radiant energy incident upon the pond is absorbed. Some of the energy is lost by reflection both from the air-water interface and from the bottom of the pond. The percentage of direct short-wave radiation reflected at the air-water interface is small except at low elevations of the Sun. It is calculable by Fresnel's reflection formula. Sverdrup (1945) gives the following figures for the percentage of total incoming radiation from sun and sky reflected from a plane water surface on a clear day at different altitudes of the sun:

Sun's altitude	..	5°	10°	20°	30°	40°	50-90°
Per cent. reflected	..	40	25	12	6	4	3

By far the major part of the daily solar radiation is received in summer in medium or low latitudes when the sun's altitude is greater than 30°, so that the air-water reflection loss of total incident short-wave radiation is not great.

Reflection of long-wave atmospheric radiation is also small, varying irregularly from 2 to 5·3 per cent. in the wavelength range of 5 to 15 μ , where atmospheric radiation is most intense (Brunt *loc. cit.*). The average value in this wavelength range is about 2 per cent. The total loss of long and short-wave radiation at the pond surface is thus calculable, provided the surface is not too violently disturbed by waves. The total loss is not great.

A larger proportion of the incident energy may be lost by reflection from the bottom of the pond, if this is highly reflecting. Before reflection at the bottom and return to the surface, much of the energy will, however, be absorbed in its passage through the water. All the long-wave atmospheric radiation penetrating the surface will be absorbed in a few mm. (Brunt 1944). In clear sea-water (which, as regards absorption of radiant energy, differs little from pure water), only 61 per cent. of solar radiation however is absorbed after penetration to a depth of 1 metre (Sverdrup 1945). The rate of absorption falls off rapidly with depth after this, so that shallow ponds, with good reflecting bottoms, may return to the air-water surface, 39 per cent. of the original short-wave radiation penetrating the surface. A second reflection will occur there for all rays of incident angle less than the critical. Calculation of the amount of energy lost by reflection from pond bottoms of various reflecting powers is obviously complicated, and in what follows it is assumed, for simplicity, that the pond bottom is a perfect absorber.

(b) The Sensible Heat Supply from the Air

If the (unknown) temperature of the surface layer of an evaporation pond is θ_w , that of the air "near" the surface θ_a , then the rate of energy influx to a small area of the pond is given by

$$\frac{dQ_s}{dt} = h(\theta_a - \theta_w), \quad \dots\dots\dots (2)$$

where h is a heat transfer coefficient. By making use of this formula, it is assumed, in accordance with the standard theory adopted in heat transfer calculations, that a very thin stagnant film of air is in contact with the liquid surface and that outside this, due to turbulence, the air temperature above the element of surface is substantially constant up to at least a moderate height above the surface. dQ_s/dt may be positive or negative, depending on whether θ_a is greater or less than θ_w . As will be seen later, for most of the diurnal cycle in shallow ponds, θ_w is less than θ_a .

When it has been necessary to use a numerical value of h , this has been obtained by use of the equation

$$h(\text{cal./cm.}^2/\text{hr./}^\circ\text{C.}) = 0.48 + 0.083v, \quad \dots\dots\dots (3)$$

where v is the wind velocity "near" the surface in ft./sec. This formula is based on data given by Fishenden and Saunders (1932) for heat transfer from plane surfaces to gases flowing parallel to them in forced convection.

In Table 1 are given examples, calculated by this formula, of the magnitude of the heat flow to 1 cm.² of pond surface at wind velocities "near" the surface of 10 and 20 ft./sec. respectively for various assumed values of $(\theta_a - \theta_w)$.

TABLE 1

Temperature Difference ($^\circ\text{C.}$)	$h (\theta_w - \theta_a)$ (cal./hr./cm. ²)	
	$v = 10$ ft./sec.	$v = 20$ ft./sec.
1	1.31	2.14
2	2.62	4.28
4	5.24	8.56
5	6.55	10.70
10	13.10	21.40

(c) Energy Losses from a Shallow Pond

The energy loss by contact with cooler air has just been considered. There remain for consideration the losses due to evaporation and to low temperature radiation.

The rate of energy loss due to evaporation dQ_E/dt follows from the Dalton equation. Where L is the latent heat per gram of the evaporating liquid, we have, per cm.² of surface,

$$\frac{Ldw}{dt} = \frac{dQ_E}{dt} = Lk(p_w - p_a), \quad \dots\dots\dots (4)$$

where k is a mass transfer coefficient, and the other symbols have their usual significance.

In this equation k , the mass transfer coefficient, may be expressed in terms of the heat transfer coefficient, h , since the laws governing heat and mass transfer between phases are essentially similar.* It is known, by combining observation and a heat balance of the wet- and dry-bulb thermometer, that, provided temperatures are expressed in °C., and vapour pressures in mm. of mercury, for water

$$\frac{h}{Lk} = 0.50, \quad \dots\dots\dots (5)$$

(see for example, Walker, Lewis, and McAdams 1927, p. 443; or Landolt and Bornstein 1943).

Equation (4) therefore becomes

$$\frac{Ldw}{dt} = \frac{dQ_E}{dt} = 2h(p_w - p_a). \quad \dots\dots\dots (6)$$

It is clear from equations (2) and (6) that the ratio of the sensible heat loss (or gain) to the heat loss by evaporation is given by the expression

$$0.5 \frac{(\theta_w - \theta_a)}{(p_w - p_a)}.$$

This is essentially Bowen's (1925) formula, and the above argument thus yields a simple derivation of it.

The magnitude of the heat loss by evaporation obviously varies very much with conditions. It is, of course, the quantity we wish to find.

(d) Radiation Losses

The long-wave radiation loss is simply given by use of Stefan's law and is $3600\sigma T^4$ hr. cm.², where σ is Boltzmann's constant and T is the (unknown) absolute temperature of the pond surface.

The magnitude of this loss at various temperatures is as follows:

Surface temperature (°C.)	..	0	10	15	20	25	30
Radiation loss (cal./cm. ² /hr.)	..	27.5	31.8	34.1	36.5	39.1	41.8
Radiation loss (cal./cm. ² /24 hr.)	..	660	763	818	876	938	1003

By combining the various rates of energy supply and loss listed above, and introducing a heat accumulation term due to changes in the sensible heat of the pond, a differential equation is obtained, in which the only unknowns are the

* This statement may not be true when large temperature lapse rates are produced by strong insolation in near-calm weather. Over water such lapse rates are rare.

pond temperature θ and its rate of change with time, $d\theta/dt$. The other variables are assumed known from meteorological records. Solution of this equation yields the pond temperature and hence the vapour pressure as a function of time. It is then possible, using the Dalton equation, to calculate evaporation rate at any time, and hence by integration to obtain the total evaporation in a period.

III. THE RATE OF EVAPORATION FROM AN ELEMENT OF POND SURFACE

Consider Figure 1, an area of 1 cm.² on the surface of a pond of depth l cm. A steady wind blows over the pond at a velocity v (ft./sec.), fixing the value of the heat transfer coefficient at some value h (cal./cm.²/°C./hr.). The air temperature just above and "near" the surface element is θ_a , and the absolute humidity p_a mm. The bottom of the pond is assumed to be black. The radiant energy of all wavelengths penetrating the surface at rate dQ_c/dt per cm.² will be completely absorbed, the strongest absorptions occurring at two sites, (a) at a thin

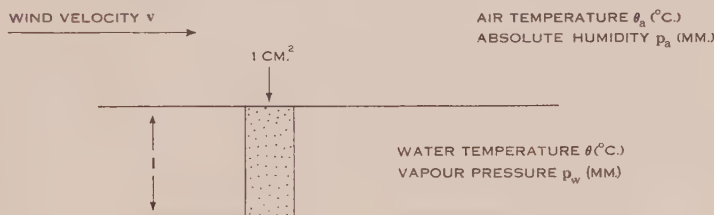


Fig. 1

surface layer, and (b) at the bottom of the pond. The resulting convection is assumed to create a uniform temperature at all points in the depth l . The specific gravity of the water is c , its specific heat ρ , its latent heat of evaporation per gram L .

Since, in the column of liquid of cross section 1 cm.² and height l cm., heat input must equal heat accumulation + heat output, the radiation input rate, dQ_c/dt , is given by

$$\frac{dQ_c}{dt} = \rho c l \frac{d\theta}{dt} + L \frac{dw}{dt} + h(\theta - \theta_a) + \frac{dQ_R}{dt},$$

where $\rho c l (d\theta/dt)$ is a heat storage term, dw/dt the weight evaporated per cm.²/hr., $h(\theta - \theta_a)$ the sensible heat loss, and dQ_R/dt the radiation heat loss in cal./cm.²/hr.

By equation (6), the evaporation term,

$$L(dw/dt) = 2h(p_w - p_a),$$

so that the above equation becomes, writing (since p_w is a known function of θ), $f(\theta)$ for p_w

$$\frac{dQ_c}{dt} - \frac{dQ_R}{dt} = \rho c l \frac{d\theta}{dt} + 2h[f(\theta) - p_a] + h(\theta - \theta_a). \quad \dots\dots\dots (7)$$

Given a knowledge of the meteorological elements, of the pond depth l , θ is the only unknown in this equation, with the exception of dQ_R/dt , the radiation loss which depends on θ . As will be shown later, θ in shallow ponds differs from

θ_a only rarely by as much as 5°C . If dQ_R/dt is calculated by assuming the pond temperature equal to the air temperature, the maximum error introduced into the net radiation absorbed is at the temperatures involved about $2.4 \text{ cal. cm.}^2 \text{ hr.}$ This error has very little effect on the general shape of the curve of $dQ_c/dt - dQ_R/dt$

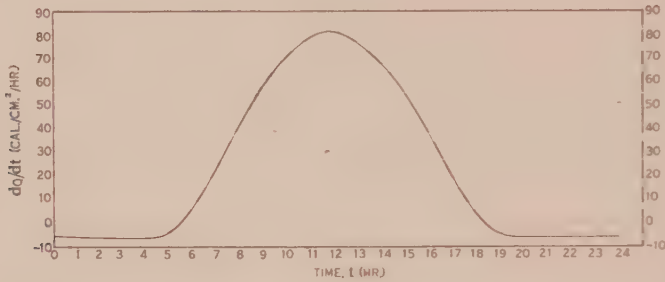


Fig. 2.—Assumed curve for net radiation absorbed through 1 cm.^2 of water surface during 24 hours, $F(t)$.

plotted against the time of day. Since, for the moment, we are concerned with the general regime of the evaporation process, we take as known $dQ_c/dt - dQ_R/dt$ as a function of time (see Fig. 2), and set the net radiation gain

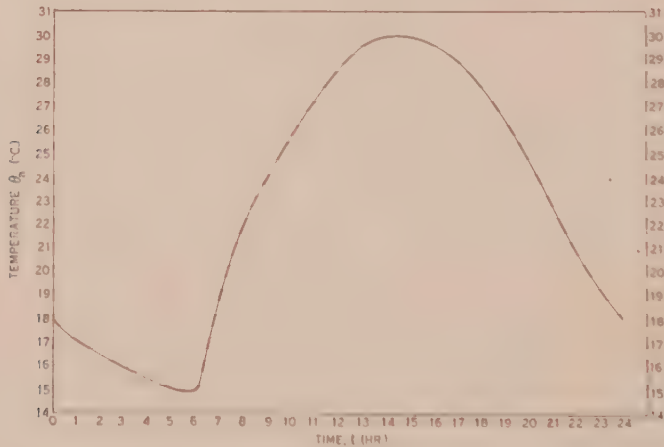


Fig. 3.—Assumed curve for diurnal variation of air temperature, $\phi(t)$.

as dQ/dt . A method of avoiding this approximation in performing practical calculations with actual measured values of the determining meteorological elements will be given later.

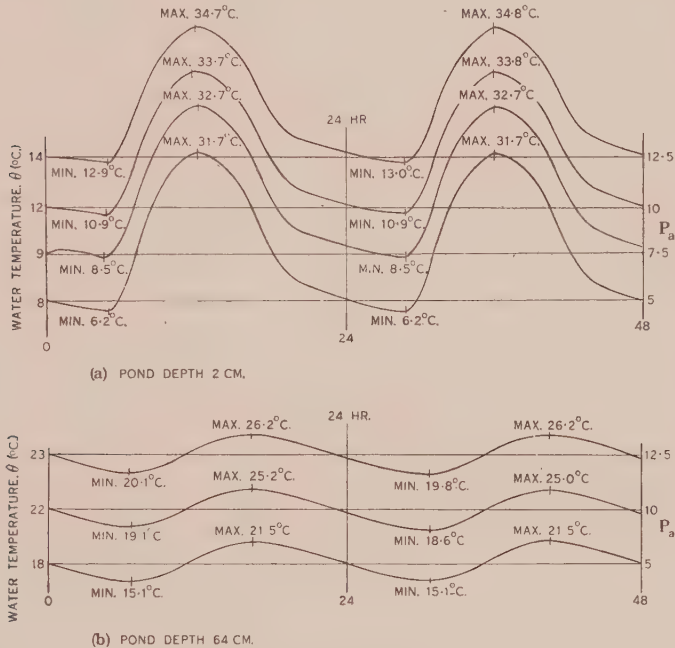
The equation then becomes

$$\frac{dQ}{dt} = \rho c l \frac{d\theta}{dt} + 2h[f(\theta) - p_a] + h(\theta - \theta_a). \quad \dots\dots\dots (8)$$

dQ/dt and θ_a are known functions of the time t , say $F(t)$ and $\varphi(t)$, respectively, and h is, for simplicity, regarded as constant. p_a is normally constant for long periods. Hence, we may write

$$F(t) = \rho c l \frac{d\theta}{dt} + 2h[f(\theta) - p_a] + h[\theta - \varphi(t)]. \quad \dots\dots\dots (9)$$

Solution of this equation yields θ and hence p_w as a known function of time t , say $\psi(t)$.



Figs. 4 (a) and 4 (b).—Water temperatures during evaporation. Differential analyser result. $h = 1.3$.

The Dalton evaporation equation is then

$$\frac{dW}{dt} = \frac{2h}{L}[(\psi)t - p_a], \quad \dots\dots\dots (10)$$

and integration of this over a time interval yields the total evaporation.

Equations (9) and (10) can be solved by various numerical methods, given $F(t)$ and $\varphi(t)$. The writer is greatly indebted to Professor D. R. Hartree, F.R.S., who, during his tenure of the chair of Theoretical Physics at Manchester University, kindly arranged for solutions to be obtained by means of the Manchester University Differential Analyser.

Since one main object of this study was to exemplify the general course of evaporation from shallow ponds under given meteorological conditions, arbitrary plots against time of day of dQ/dt , the net radiation absorbed, and of the air temperature θ_a were constructed. These are shown in Figures 2 and 3. They

were intended to represent roughly these meteorological conditions during cloudless weather in summer in, say, South Australia. These are the functions $F(t)$ and $\varphi(t)$. Suitably combined with a plot of the vapour pressure of water, they were fed into the input of the machine. The output drew a graph of θ as a function of time, i.e. a solution of the equation. Twenty-nine solutions of the equation applied to water evaporation were obtained for various steady values of the parameters, pond depth l , absolute humidity p_a , and heat transfer coefficient h .

To solve equation (9), an initial condition is necessary—the original water temperature. This was estimated, and the analyser run until the transients died out, and a periodic curve obtained. Only a few cycles were found necessary.

Typical curves of pond temperature, as drawn by the analyser, are reproduced in Figures 4 (*a*) and 4 (*b*). Examples of the cumulative evaporation curves as drawn by the analyser are not reproduced. Neither are the more interesting plots derivable from those of evaporation rates at different times in the diurnal cycle, since these are of the same general shape as the temperature curves, though more peaked at the maxima. Quantitative data obtained from all the solutions are given in Tables 2 and 3.

Examination of the curves of Figures 2 and 3 and of Tables 2 and 3 indicates some points of interest. As would be expected, the curves show that the evaporation rates from the shallowest ponds vary in magnitude during the diurnal cycle, much more than do those from the deepest, the shallow ponds showing a higher maximum and a lower minimum temperature and hence evaporation rate than the deeper ponds. These differences are, of course, due to the different degrees of damping introduced by the differing heat storage terms. The total amount of evaporation per day is in all cases greater the shallower the pond. Unlike the differences between maximum and minimum rates, this effect is not very marked, the evaporation from ponds of 2 cm. depth being only 5–7 per cent. greater than that from ponds 64 cm. deep. The calculations refer to ponds with completely absorbing and relatively non-conducting bottoms. If any marked reflection occurs at the bottom, shallow ponds will in practice yield less total evaporation than deep ponds exposed to the same conditions. The slightly greater evaporation from shallow ponds is due to such ponds receiving, under the conditions specified, a greater inflow of heat from the overlying air. This is clear from the values of the mean pond temperature, which are lower the shallower the pond, so that $h(\theta_a - \theta)$ is greater.

A similar explanation holds for the fact that the total evaporation is greater the lower the absolute humidity. The lower the latter, the lower is the mean pond temperature, and again, the greater the contribution of sensible heat from the air to the total energy used in evaporation, or the smaller the loss of energy in this way.

Cummings and Richardson (1930) have suggested that evaporation can be calculated merely from a knowledge of the total radiation input. Tables 2 and 3 show that considerable errors would occur if this procedure were adopted. The evaporations calculated here are from ponds which each absorb per day the

same amount of radiation energy—558 cal./cm.². The daily evaporation to be expected if the radiation energy absorbed were the sole source of supply and if it were completely converted to latent heat of evaporation, is obtained by dividing the daily energy absorbed by the latent heat of the water at the water temperature, viz. 585 cal./g. The evaporation thus calculated amounts to

TABLE 2

EVAPORATION OF WATER

Meteorological conditions: $h=1.3$; total radiation absorbed 588 cal./cm.²/day; max. air temp. 30 °C.; min. air temp. 15 °C.; range 15 °C.; mean air temp. θ_a 22.7 °C.

P_a (mm.)		Pond Depth (cm.)						Evaporation by Formula (13) (cm./day)
		2	4	8	16	32	64	
5.0	θ max. ..	31.5	31.6	30.4	28.2	24.5	21.7	1.19
	θ min. ..	6.2	6.8	7.4	9.6	12.7	15.3	
	θ range ..	25.3	24.8	23.0	18.6	11.8	6.4	
	θ mean ..	17.0	17.5	17.5	18.3	18.3	18.3	
	Evap. (cm./day)	1.27	1.26	1.24	1.21	1.21	1.19	
7.5	θ max. ..	32.6	32.2	31.6	29.4	26.1	—	1.11
	θ min. ..	8.5	9.0	9.6	11.2	14.5	—	
	θ range ..	24.1	23.2	22.0	18.2	11.6	—	
	θ mean ..	18.7	18.6	19.3	19.6	20.1	—	
	Evap. (cm./day)	1.17	1.17	1.15	1.14	1.12	—	
10.0	θ max. ..	33.6	33.6	32.8	30.8	27.4	24.9	1.02
	θ min. ..	10.9	11.2	11.8	13.2	15.8	18.6	
	θ range ..	22.7	22.4	21.0	17.6	11.6	6.3	
	θ mean ..	20.4	20.8	20.9	21.4	21.7	21.8	
	Evap. (cm./day)	1.09	1.08	1.07	1.04	1.03	1.03	
12.5	θ max. ..	34.6	34.6	33.8	32.2	28.8	26.2	0.95
	θ min. ..	13.1	13.2	13.6	15.2	17.4	20.0	
	θ range ..	21.5	21.4	20.2	17.0	11.4	6.2	
	θ mean ..	21.9	21.9	22.1	22.9	23.2	23.1	
	Evap. (cm./day)	1.01	1.00	0.98	0.98	0.95	0.95	

0.95 cm. Even in the case of the deeper ponds, our calculated evaporations vary, according to conditions from 1.19 to 0.84 cm., i.e. from 25 per cent. above to 17 per cent. below the figure deduced solely from the influx of radiant energy. The conditions specified here are conditions of intense insolation. With a smaller influx of radiant energy, still greater divergencies are to be expected.

From the practical point of view of predicting evaporation rates, a more important aspect of the results is indicated by the last column in Tables 2 and 3.

The evaporations listed in this column were calculated from an approximate formula, using the mean air temperature and the value of the total radiant energy received per day.

It will be noted that the approximate calculation yields almost the same result as the exact formula, for ponds exceeding 16 cm. in depth.

TABLE 3
EVAPORATION OF WATER
Meteorological conditions: $h=0.65$; others as shown in Table 2

p_a (mm.)		Pond Depth (cm.)		Evaporation by Formula (13) (cm./day)
		4	64	
5	θ max.	41.2	26.8	0.92
	θ min.	5.2	20.9	
	θ range	36.0	5.9	
	θ mean	20.9	23.9	
	Evap. (cm./day)	1.02	0.92	
10	θ max.	42.8	29.6	0.85
	θ min.	9.8	23.8	
	θ range	33.0	5.8	
	θ mean	23.8	26.8	
	Evap. (cm./day)	0.94	0.88	
12.5	θ max.	43.6	30.8	0.83
	θ min.	11.8	25.0	
	θ range	31.8	5.8	
	θ mean	25.0	27.8	
	Evap. (cm./day)	0.90	0.84	

The approximate formula is obtained from equation (8) merely by omitting the heat storage term $\rho c l (d\theta/dt)$, and by using mean daily values of the air temperature θ_a , and of the total radiation absorbed per day Q . In deriving it, as in deriving the more exact formula, a constant wind velocity, i.e. a constant " h ", has, for simplicity, been postulated.

Equation (8) then becomes

$$Q = 2h(p_w - p_a) + h(\theta_w - \theta_a). \quad (11)$$

Rearranging

$$2p_w + \theta_w = \frac{Q}{h} + 2p_a + \theta_a. \quad (12)$$

The term $2p_w + \theta_w$ is thus found in terms of known variables, and it is a simple matter to find from a table of water vapour pressure *versus* temperature,

the unique value of θ_w or of p_w which corresponds to any value of $2p_w + \theta_w$. (It is clear that the temperature of ponds containing liquids of lower vapour pressure than water (e.g. brine) will be higher under the same conditions than that of water ponds.)

Writing

$$p_w = f'(2p_w + \theta),$$

$$p_w = f'\left(\frac{Q}{h} + 2p_a + \theta_a\right).$$

Since the total depth of liquid evaporated in unit time, E , is

$$E = \frac{2h}{cL}(p_w - p_a),$$

we have

$$E = \frac{2h}{cL}\left[f'\left(\frac{Q}{h} + 2p_a + \theta_a\right) - p_a\right]. \quad \dots\dots\dots (13)$$

This formula gives the evaporation in cm./hr., when Q is the mean radiation in calories absorbed per cm.²/hr., and h is expressed in cal./cm./hr./°C.

This equation is readily transformed to Penman's (1948) equation (16).

For, let p_e be the vapour pressure of water at the air temperature θ_a , and let Δ be the slope of the vapour pressure curve at θ_a .

Making Penman's assumption of the approximate constancy of Δ over the range θ_w to θ_a ,

$$\theta_w - \theta_a = \frac{p_w}{\Delta} - \frac{p_e}{\Delta},$$

and equation (11) becomes

$$Q = 2hp_w + \frac{hp_w}{\Delta} - 2hp_a - \frac{hp_e}{\Delta},$$

from which

$$p_w = \frac{Q + 2hp_a + hp_e/\Delta}{2h + h/\Delta},$$

and the evaporation $E = 2h(p_w - p_a)$ is then

$$E = \frac{Q + (p_e - p_a)h/\Delta}{1 + 0.5/\Delta},$$

which, in the present notation, is Penman's equation.

It may be noted that Δ is 1.1 mm. per degree at 20 °C. and 1.43 at 25 °C.

IV. DISCUSSION

Equation (13) has been checked against observation in South Australia by Bonython (1948, 1950) who, during 1947-51, determined measured evaporation rate in tank evaporimeters and at the same time calculated evaporation rate from the necessary meteorological data using equation (13). The radiant energy influx was determined by means of a Kipp and Zonen solarimetric thermopile, fitted with a Cambridge Type B millivolt thread recorder. The thermopile was standardized against a Smithsonian Silver Disc Pyrheliometer. Some of these

observations have been published (Bonython 1950). Seasonal curves of the measured and calculated rates showed good correlation, although there was a discrepancy between the respective absolute values of the rates. Work during 1950-51 (unpublished data) eliminated a previously-neglected source of error in measured evaporation, viz. sensible heat exchange between buried evaporimeter and surrounding earth—reducing the average discrepancy between measured and calculated evaporation to less than 10 per cent. Penman also reports the results of many evaporation observations which agree with the predictions of his equation, which equation, as just shown, is, apart from its approximation, equivalent to equation (13), tested by Bonython.

Provided the radiant energy input is known accurately, a fair measure of agreement between the observed evaporation in climates with high insolation and that predicted by these two equations might be expected, since the radiant energy term in the heat balance is much greater than the term representing sensible heat gain or loss. Nevertheless, we have seen that even under such conditions the maximum evaporation calculated here is 43 per cent. greater than the minimum. The difference is entirely ascribable to the variations in the sensible heat gain or loss, determined by different humidity and wind conditions.

It is important to note the assumptions used in deriving the evaporation equation, and the limitations that these may impose on its use. The most serious limitations and assumptions are (1) that the equation purports to give the evaporation from an element of surface, and not from a large area, (2) that the surface temperature of the water is also that of the main water bulk. Difficulties also arise in determining what is meant by wind velocity "near" the surface. In Bonython's (1950) check, wind speed was measured by anemometer at 3 ft. above the surface. Instead of finding h from Table 1 data he derived it from the empirically-determined water vapour diffusion coefficients, related to the wind at 3 ft., for the tank evaporimeters concerned. (Values of h so found agree closely with those of Table 1 when it is assumed that x of Table 1 is that at the 3 ft. level.)

As regards the first limitation, it is clear that over a surface of any extent, the absolute humidity of the atmosphere over and "near" the water surface will rise as we proceed downwind, unless possibly atmospheric turbulence is very great indeed. Assuming a uniform surface temperature, Jeffreys (1918), Giblett (1921), and Sutton (1934) have given treatments of this problem.

The assumption of a uniform water temperature in the surface and the main bulk is unlikely to hold good under night conditions, but as evaporation is least at night, it may not be very important.

An objection which holds against the application of equations derived here, and also against Penman's derivation, is the neglect of the difference between water and air temperature in determining the radiation loss from the pond.

This can be surmounted for the approximate equation in the following way, for the derivation of which the writer is indebted to Mr. A. Callender.

Equation (11) may be written

$$Q_P - Q_R + Q_L = 2h(p_w - p_a) + h(\theta_w - \theta_a). \quad (14)$$

Here Q_P is the short-wave radiation absorbed, Q_R the return long-wave radiation from the pond, and Q_L the long-wave radiation from the atmosphere.

Where T_w and T_a are the absolute water and air temperatures and x is Brunt's coefficient equation [(1)], (14) becomes

$$Q_P - 3600\sigma T_w^4 + 3600\sigma T_a^4 x = 2h(p_w - p_a) + h(\theta_w - \theta_a). \quad (15)$$

The long-wave radiation terms in (15) can be combined into the expression

$$3600\sigma(T_w^4 - T_a^4 + T_a^4 - T_a^4 x).$$

The part in brackets can be rewritten as

$$T_a^4 \left\{ \left(\frac{T_w}{T_a} \right)^4 - 1 \right\} + T_a^4 (1 - x),$$

or

$$T_a^4 \left[\left(1 + \frac{T_w - T_a}{T_a} \right)^4 - 1 \right] + T_a^4 (1 - x).$$

Expanding the first term by the binomial theorem, and neglecting the second term of the expansion $(T_w - T_a/T_a)^2$ compared with unity, we obtain

$$4T_a^3(T_w - T_a) + T_a^4(1 - x).$$

Since $T_w - T_a = \theta_w - \theta_a$, equation (15) becomes

$$Q_P - 3600\sigma T_a^4(1 - x) = 2h(p_w - p_a) + (h + 14,400\sigma T_a^3)(\theta_w - \theta_a), \quad (16)$$

in which θ_w and, of course, its dependent variable p_w are the only unknowns.

The left-hand side of this equation is the term Q of equation (11), i.e. the net radiation absorbed calculated assuming equality of air and water temperatures.

Equation (11) when corrected therefore becomes

$$Q = 2h(p_w - p_a) + (h + r)(\theta_w - \theta_a), \quad (17)$$

where $r = 14,400\sigma T_a^3$, and the calculation of the water temperature is effected by the equation

$$2p_w + \theta_w \left(1 + \frac{r}{h} \right) = \frac{Q}{h} + 2p_a + \theta_a \left(1 + \frac{r}{h} \right). \quad (18)$$

Neglect of the exact radiation correction thus yields a value of $2p_w + \theta_w$ in equation (12), which is in error by $(\theta_w - \theta_a) \frac{r}{h}$. At 20 °C. r has the value 0.5.

In conclusion, and to sum up, the equations given in this paper purport to provide a method of calculating the mean water temperature of shallow ponds of depth greater than say 30 cm., over short periods, provided the air temperature, the absolute humidity, the wind velocity and its effect on the heat transfer coefficient, and the solar radiation absorbed are known. If no instrumental records of the latter quantity are available, the solar radiation received over short periods is calculable approximately for any latitude and time of the year,

from a knowledge of the daily hours of sunshine, the cloudiness, the air temperature, and the humidity. The necessary formulae are to be found in Brunt's (1944) treatise or Penman's (1948) paper. From a knowledge of the water temperature, the absolute humidity of the air, and the heat transfer coefficient, the depth of water evaporated in short periods is then readily determined by the Dalton equation.

V. ACKNOWLEDGMENTS

Very great thanks are due to Professor D. R. Hartree for making available the Manchester University Differential Analyser for the solution of the differential equation; to his staff for operating it; and to Mr. A. Callender for suggesting its use and for helpful discussion during the preparation of this paper.

VI. REFERENCES

- BONYTHON, C. W. (1948).—*Aust. J. Instrum. Technol.* **4**: 209.
BONYTHON, C. W. (1950).—*Trans. Roy. Soc. S. Aust.* **73**: 198.
BOWEN, T. S. (1925).—*Phys. Rev.* **27**: 779.
BRUNT, D. (1944).—"Physical and Dynamical Meteorology." (Cambridge Univ. Press.)
CUMMINGS, N. W. (1921).—*J. Elect. West. Ind.* **46**: 491.
CUMMINGS, N. W., and RICHARDSON, B. (1930).—*Phys. Rev.* **30**: 527.
FISHENDEN, M., and SAUNDERS, O. A. (1932).—"The Calculation of Heat Transmission." (H.M.S.O.: London.)
GIBLETT, M. A. (1921).—*Proc. Roy. Soc. A* **99**: 472.
JEFFREYS, H. (1918).—*Phil. Mag.* **35**: 270.
KIMBALL, H. H. (1928).—*Mon. Weath. Rev.* **56**: 393.
LANDOLT, H. H., and BORNSTEIN, R. (1943).—"Physikalische Tabellen." 5th Ed. Vol. 2. p. 1364. (Edwards: Ann Arbor, Michigan.)
PENMAN, H. L. (1948).—*Proc. Roy. Soc. A* **193**: 120.
SUTTON, O. G. (1934).—*Proc. Roy. Soc. A* **146**: 701.
SVERDRUP, H. U. (1945).—"Oceanography for Meteorologists." (Prentice-Hall: New York.)
WALKER, W. H., LEWIS, W. K., and McADAMS, W. H. (1927).—"Principles of Chemical Engineering." 2nd Ed. (McGraw-Hill: New York.)

THE SPONTANEOUS EMULSIFICATION OF MIXTURES OF OLEIC ACID AND PARAFFIN OIL IN ALKALINE SOLUTIONS

By W. W. MANSFIELD*

[*Manuscript received January 21, 1952*]

Summary

Interfacial tension values occurring during the transfer of oleic acid from paraffin oil to aqueous alkali have been determined by the sessile-drop method.

It is postulated that the ease of emulsification observed with these systems arises from the nature of the changes in interfacial tension associated with the transfer.

I. INTRODUCTION

The ease of emulsification of fatty acid and oil mixtures in alkaline solution is well known, and is utilized in the preparation of emulsions. The phenomenon of spontaneous emulsification was studied extensively by McBain and Woo (1937); these authors gave a review of early work on the subject. They found that a deep layer of "Nujol" containing 5 or 10 per cent. lauric acid placed carefully on a dilute alkaline solution emulsified slowly but spontaneously. Subsequent analysis of the two phases revealed that this process had been accompanied by the transfer of lauric acid from the oil to the aqueous phase. Sodium hydroxide concentrations above 0.005N stopped emulsification.

A second and much more rapid type of spontaneous emulsification is obtained if the oil is present as a small lens on the surface of the aqueous phase. For such systems the emulsification is considered to be due to local movements, which will not be retarded by gravitation, resulting from inequalities in interfacial tensions over the interface (McBain and Woo 1937). In recent years such emulsification has been observed by Kling and Schwerdtner (1941) for droplets of oleic acid in certain detergent solutions, by Pospelova and Rehbinder (1942) for hydrocarbon-water systems containing soaps in each phase, and by Schulman and Cockbain (1940) for mixtures of cholesterol and "Nujol" in the presence of sodium cetyl sulphate solutions.

Stackelberg, Klockner, and Mohrhauer (1949) obtained evidence suggesting that the spontaneous emulsification of oleic acid-paraffin oil mixtures in dilute alkali results from the existence of a negative interfacial tension at the oil-water interface. The interfacial tension of a 10 per cent. oleic acid mixture decreased rapidly as the pH of the aqueous solution was increased; a line drawn through the experimental points suggested that a zero interfacial tension existed at about pH 8.7. Using the drop-volume method these authors found interfacial tension measurements impossible between pH 9 and 12 because of the ease of emulsification of the oil. Above pH 12, the interfacial tension increased with pH.

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

The nature of the variation of interfacial tension with pH suggested the existence of negative interfacial tension values between pH 9 and 12. A study of their measurements, after applying the corrections necessary when using the drop-weight method (Harkins and Brown 1919), suggests that oils containing more than about 6 per cent. of oleic acid would show negative interfacial tensions against an alkaline solution.

Stackelberg, Klockner, and Mohrhauer (1949) showed also that oleic acid is transferred from the oil to the aqueous phase between pH 9 and 12. This

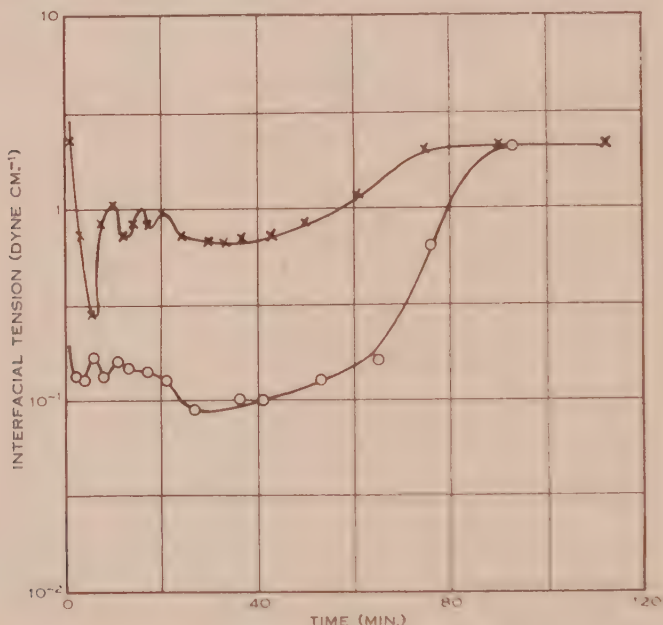


Fig. 1.—The variation of interfacial tension with time for system *B* (Table 1). Initial phases were 0.001N sodium hydroxide v. 0.0078M oleic acid in "Nujol". Aqueous solution was stirred (x) or unstirred (o).

transfer leads to concentration conditions quite different from those existing outside this pH range. On this ground extrapolation appears unjustified.

In the present work, interfacial tensions have been measured during the transfer. In the initial stages small but positive values were recorded. It is shown that the ease of emulsification is associated more with the change in interfacial tension than with the actual values.

II. THE MEASUREMENT OF LOW INTERFACIAL TENSIONS

The changes of interfacial tension, associated with the transfer of oleic acid from a paraffin oil drop to a surrounding alkaline solution, were measured by the sessile-drop method.

A drop of oil of volume about 0.005 ml. was placed on the underside of a small glass platform supported in a 40 ml. optical cell. A V shaped support for the platform consisted of two similar glass plates each resting against the

other at the bottom of the cell and also against a side wall of the cell. The resting edges of these plates were ground parallel to each other to 1 part in 10,000. Usually the stability of sessile drops is ensured by the use of a slightly hollowed plate. This was not feasible because the oil droplets were almost flat for many measurements.

The optical cell was placed on a mechanical stage. A pin with a finely ground point was mounted upright on the stage for use as a reference point, and both drives of the stage were checked so that the plate to which the drop adhered was shown to be horizontal with a precision approaching 1 part in 10,000. This proved sufficient to maintain stability of the sessile drops studied.

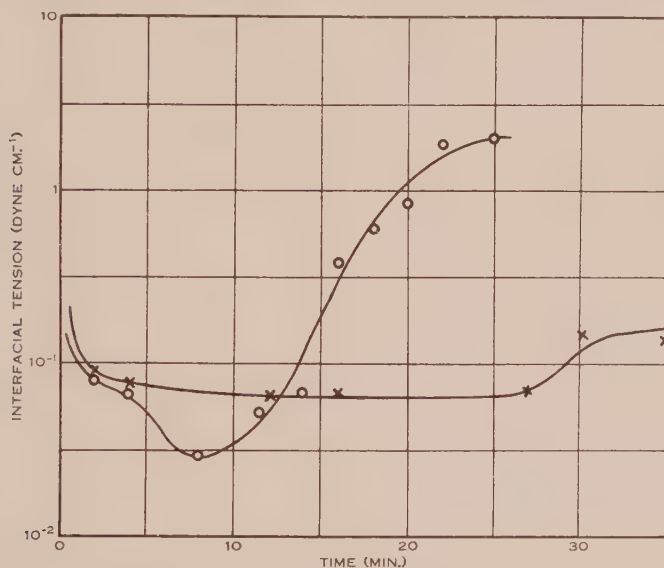


Fig. 2.—The variation of interfacial tension with time for system *F* (X) and system *G* (O) (see Table I).

A drop was then viewed through a low-power microscope fitted with cross-hairs and mounted on a cathetometer. The image of one edge of the drop at its contacting plane was set at the cross-hairs; this edge was well defined because the plate to which the drop adhered acted as a mirror. Using the drive of the mechanical stage, the other edge of the drop was brought into view. The plate was considered horizontal when this image coincided with the cross-hairs. Any change in position was obtained by tapping gently the appropriate member of the V-shaped support.

After alignment of the plate, 20 ml. of freshly-boiled distilled water was added. Drops were displaced by the aqueous solution unless the plate previously had been wetted and well rubbed with the oil. After this treatment a drop would adhere with a contact angle of about 150° (measured in the oil phase).

Following addition of the water, a gentle stream of carbon dioxide-free air was introduced a little above the water surface. A small and appropriate amount of a stock alkaline solution was added, and the solution stirred carefully for about

10 seconds. Tests with indicator solutions showed this to be sufficient for complete mixing. A stop-watch was started on the addition of the alkali.

The measurements made to allow estimation of interfacial tensions were h cm., the vertical distance between the equatorial plane and the bottom of a drop, and l cm., the radius of the drop at the equator. The formula used was derived by Worthington in 1885 and is discussed by Burdon (1949).

It reads

$$\gamma = \frac{1}{2} \Delta d \cdot g \cdot h^2 \cdot \frac{1 \cdot 641 l}{1 \cdot 641 l - h},$$

where Δd g. per ml. is the difference in density between the two phases, and γ dyne per cm. is the boundary tension.

Comparison with calculations made from the tables of Bashforth and Adams (1883) showed that this formula leads to increasing errors when l/h is less than 2.5. Accordingly the tables of Bashforth and Adams were used in this range.

Values of h were measured with a microcathetometer reading to 10^{-4} cm. Drop radii were measured, using the mechanical stage, to 0.01 cm. The density difference between the two phases was assumed independent of solute concentrations. All measurements were made at $20 \pm 0.5^\circ \text{C}$.

TABLE I
APPROXIMATE STEADY-STATE (FIG. 2), LOWEST OBSERVED, AND EQUILIBRIUM VALUES OF INTERFACIAL TENSION OCCURRING DURING THE TRANSFER OF OLEIC ACID FROM PARAFFIN OIL TO AQUEOUS ALKALI AT 20°C .

System	Initial Molality of Oleic Acid in Oil	Initial pH of Solution	Ionic Strength of Solution	Interfacial Tension (dyne cm. ⁻¹)			Time Required for Equilibrium (min.)
				Approx. Steady-State	Lowest Observed	Equilibrium	
A	0.0040	10.9	0.001	0.07	0.038	3.0	60
B	0.0078	10.9	0.001	0.10	0.080	2.2	80
C	0.0193	10.9	0.001	0.03	0.029	—	> 110
D	0.035	10.9	0.001	0.1	0.059	—	> 140
E	0.051	10.9	0.001	0.08	0.057	—	> 140
F	0.0078	10.8	0.34	0.07	0.064	0.15	30
G	0.0078	13.0	0.10	0.06	0.029	1.9	25
H	0.035	13.0	0.10	0.03	0.026	—	—
J	0.035	9.7	1.09	0.05	0.041	—	—

The drop volume used was chosen to give accurate measurements of interfacial tension below 0.1 dyne per cm., where the determinations may be in error by as much as 3 per cent. Between 0.1 and about 2.5 dynes per cm., the results may be subject to errors approaching 10 per cent.

A number of the plots obtained are recorded in Figures 1 and 2. These and other results are summarized in Table 1.

III. DISCUSSION

It has been demonstrated (Stackelberg, Klockner, and Mohrhauer 1949) that oleic acid is transferred almost completely from paraffin oil to a surrounding alkaline solution between the approximate pH limits of 9 and 12. It is apparent from the data obtained for systems *A* to *E* that interfacial tensions are low during the transfer. As equilibrium is approached the low residual oleic acid content of the oil phase leads to a relatively high equilibrium interfacial tension.

It was shown also by the above authors that the transfer of fatty acid is hindered as the pH of the alkaline solution is increased above 12; almost no acid migrates into a solution of pH 14. It was suggested that the effect follows from the decreased solubility of sodium oleate in concentrated alkaline solutions. Data reported in this paper do not confirm this effect at high pH value. The equilibrium interfacial tension observed by using a solution of pH 13 (system *G*) was only slightly smaller than that found with a solution of pH 10.9 (system *B*), and the time required for equilibrium to be reached was very much less. It is considered that the discrepancy arises from the widely-differing oil-solution ratios used by the writer and the above authors. For this work the volume ratio was 1/6000; whereas the distribution data of Stackelberg, Klockner, and Mohrhauer (1949) were obtained analytically after centrifugal separation of concentrated emulsions. In the latter study the observation of any effect from a decreased solubility of sodium oleate is more probable.

A marked salt effect was observed, however, on the introduction of 0.34*N* sodium chloride into an aqueous phase containing 0.001*N* sodium hydroxide (system *F*). For system *F*, the equilibrium interfacial tension of 0.15 dyne cm.¹ indicated that the transfer of fatty acid had been decreased considerably. The salting out of soap by salt solutions is well known.

Spontaneous emulsification for these systems cannot be due to the existence of a negative interfacial tension, as the values recorded for the systems *C*, *D*, *E*, and *J* are positive. The lowest recorded values of interfacial tension were obtained for the systems *H*, *G*, and *C*. Systems *H* and *G* do not emulsify as readily as any of the other systems, with perhaps the exception of system *A*. It follows that low interfacial tension values, in themselves, do not lead to ease of emulsification.

It will be observed that the plots of Figure 1 have not been drawn as smoothed curves. This was done because oscillations in drop shape were observed readily while making the measurements. Owing to the rapidity of the oscillations compared with the time involved in measuring the two values required for the calculation of interfacial tension, the line drawn through the experimental points corresponds only approximately to the actual changes in interfacial tension. While the mean values of interfacial tension were reproducible from drop to drop, the period and magnitude of the oscillations were not. On a few occasions oscillations corresponding to an interfacial tension change of about 2 dynes per cm. were observed.

The transfer of fatty acid from oil to alkaline solution involves four processes: diffusion of fatty acid to the oil-solution interface, migration of acid

through the interface, ionization of the acid, and diffusion of the oleate ion and its ionic atmosphere into the bulk of the solution. The ionization is assumed to be rapid. In Figure 1 it is shown that continuous mechanical stirring (1 rev. per sec.) of the aqueous phase reduced the time required for the transfer of oleic acid by about 20 per cent. Stirring minimizes diffusion within the aqueous phase as a factor influencing the rate of transfer. The reduction in the time required for complete transfer is sufficiently small to indicate that diffusion within the aqueous phase is several times more rapid than are the combined rates of diffusion through the oil and migration through the interface.

The existence of oscillations in interfacial tension during the transfer of fatty acid points to the conclusion that diffusion within the oil phase is slower than migration through the interface. Thus a drop having a low interfacial tension distorts readily. Any portion of this distorted drop which is thinner than another will be more rapidly depleted of its oleic acid since there is less acid underneath each unit of area exposed to the aqueous solution. The high viscosity of the paraffin oil hinders both diffusion and mixing so that transport of oleic acid to regions of low concentrations will be slow. At those parts of the surface (over the thin sections) where the concentration of oleic acid is small the interfacial tension will rapidly rise. Thus the variable thickness of the drop and the high viscosity of oil produce differing interfacial tensions in sections of the drop. These differences are sufficient to cause movement of oil which extinguishes the differences in interfacial tension and produces emulsification of pendent drops. With carefully aligned sessile drops oscillations of the surface are produced but there is no emulsification. If the plate below which a sessile drop is held is not quite horizontal, emulsification occurs in the direction of slope of the plate.

When a drop of 5 per cent. oleic acid in paraffin oil was held on the *upper* surface of a hydrophobic glass plate and surrounded with 0.001N sodium hydroxide, apparently random emulsification occurred. Close examination of the process suggested that the emulsification droplets were formed at the end of long streamers of oil. By using an oil containing 2 per cent. of oleic acid, the phenomenon could be observed clearly. It is illustrated by Plate I. On addition of alkali, long streamers of oil rose into the solution, indicating the fall in boundary tension. During an interval of a few seconds, a spherical drop of oil formed at and parted from the narrow tip of each streamer. After the detachment of the droplet, the thin column of oil broadened and subsided. Both of these processes are consistent only with an increased interfacial tension. The column of oil then lengthened again and the process was repeated.

A consequence of this mechanism is that transfer of oleic acid is required to obtain varying interfacial tensions which cause spontaneous emulsification. When a drop of an oleic acid paraffin oil mixture is surrounded by a concentrated alkaline solution or a solution of high ionic strength, transfer of fatty acid is diminished. The processes automatically resulting from the transfer are dampened, and unassisted emulsification proceeds less readily or not at all.

The slow transport of oleic acid within the oil phase has been postulated to be of major importance in the outlined mechanism of emulsification. It is

significant that solutions of oleic acid in solvents such as benzene do not emulsify as readily as do solutions in paraffin oil. To obtain a finely-divided emulsion of solvents of low viscosity and consequently high diffusion coefficient some form of stirring is necessary. Similarly, the readiness with which oleic acid-paraffin oil mixtures emulsify spontaneously in alkali decreases with increasing temperature (Stackelberg, Klockner, and Mohrhauer 1949). It is desirable, however, to show that the rate of transfer of oleic acid from paraffin oil to aqueous alkali is determined mainly by the diffusion of fatty acid within the oil.

The partition coefficient of oleic acid between paraffin oil and a very dilute solution of sodium oleate is about 10^5 , and the hydrolysis constant of the solution is about 1.2×10^{-9} (Mansfield, unpublished data). For the experimental results considered in Figure 1, the ratio of the volumes of paraffin oil and of the alkaline solution of pH 10.9 was about 1/6000. From these figures, the concentration of oil present in the drop at equilibrium was calculated to be about 1/10,000 of that originally present.

The diffusion coefficient of oleic acid in paraffin oil, estimated from the Sutherland-Einstein equation (Alexander and Johnson 1949), was about 10^{-6} cm.² sec.⁻¹. In dilute soap solutions the diffusion coefficient of sodium oleate approaches 8×10^{-6} cm.² sec.⁻¹ (McBain 1933); hence diffusion of sodium oleate away from the interface should not be important for the rate of transfer of oleic acid.

Accordingly it was considered that the transfer could be represented approximately by the diffusion of oleic acid from a slab of suitable dimensions into a solution always at zero concentration. The problem is analogous to a heat transfer problem treated by Carslaw and Jaeger (1948). Using the equation given by these authors on pp. 257-8, it was calculated that transfer of oleic acid should be virtually complete after about 5000 sec. provided an activation energy of about 18 RT was required by an oleic acid molecule in order to pass through the interface. The order of the activation energy is reasonable. Hence the rate of transfer of oleic acid appears largely to be determined by the diffusion coefficient of oleic acid in paraffin oil.

The type of slow spontaneous emulsification discussed by McBain and Woo (1937) was observed in this work during measurement of interfacial tensions. Particularly with the more concentrated oleic acid-oil mixtures, a turbid layer formed around the drops as the oleic acid was transferred. Since the equilibrium concentration of sodium oleate formed after the transfer was very low, the oil removed in this way did not remain emulsified, but collected as a film on the surface of the solution.

IV. ACKNOWLEDGMENT

The advice and interest of Dr. K. L. Sutherland were invaluable in the course of this investigation.

V. REFERENCES

- ALEXANDER, A. E., and JOHNSON, P. (1949).—"Colloid Science." Vol. 1. p. 259. (Clarendon Press: Oxford.)
BASHFORTH, F., and ADAMS, J. C. (1883).—"An Attempt to Test the Theories of Capillary Action." (Cambridge Univ. Press.)

- BURDON, R. S. (1949).—"Surface Tension and the Spreading of Liquids." 2nd Ed. pp. 15-18.
(Cambridge Univ. Press.)
- CARSLAW, H. S., and JAEGER, J. C. (1948). "Conduction of Heat in Solids." pp. 257-8.
(Clarendon Press: Oxford.)
- HARKINS, W. D., and BROWN, F. E. (1919).—*J. Amer. Chem. Soc.* **41**: 499-524.
- KLING, W., and SCHWERDTNER, H. (1941).—*Melliand Textilber.* **22**: 21-8.
- McBAIN, J. W., and WOO, TS-MING (1937).—*Proc. Roy. Soc. A* **163**: 182-8.
- McBAIN, M. E. (1933).—*J. Amer. Chem. Soc.* **55**: 545-51.
- POSPELOVA, K., and REHBINDER, P. (1942).—*Acta Physicochim. U.R.S.S.* **16**: 71-87.
- SCHULMAN, J. H., and COCKBAIN, E. G. (1940). *Trans. Faraday Soc.* **36**: 651-61.
- STACKELBERG, M. V., KLOCKNER, E., and MOHRHAUER, P. (1949).—*Kolloidzchr.* **115**: 53-61.

SPONTANEOUS EMULSIFICATION



Spontaneous emulsification of oil containing 2 per cent. oleic acid. $\times 14$.

CONJUGATION ENERGY

By R. D. BROWN*

[*Manuscript received December 12, 1951*]

Summary

The relationship between conjugation energies and self-atom-polarizabilities, previously derived from certain approximate theoretical relationships, is obtained in a more accurate form by an empirical least-squares calculation. An analogous relationship is also derived from calculating conjugation energies when the overlap integral is taken into account. A tabulation of all relevant polarizability data from the literature is presented, and a method of extending the data to larger molecules is described and exemplified. A possible alternative method of deriving conjugation energies is briefly discussed.

I. INTRODUCTION

In a paper concerned with bond localization energies (Brown 1949) an empirical equation was presented for computing the LCAO conjugation energy, neglecting overlap. The use of conjugation energies as a means of determining resonance energies has been found to be so convenient that in the present paper the empirical relation is studied in more detail, and in particular the corresponding equation is derived for the conjugation energy with overlap included. The present results have already been utilized for computations of various resonance energies involved in the localization theory of bond reactions (Brown 1950*a*) and the Diels-Alder reaction (Brown 1950*b*, 1950*c*).

II. DEFINITION AND FORMULAE

The conjugation energy, i.e. the energy† change accompanying the conjugation of two initially separated conjugated systems A and B is defined as the increase in the resonance energy, R , accompanying this linking process. Thus

$$C_{ab} = R_{AB} - R_A - R_B,$$

where the subscripts a and b indicate that the conjugation energy depends upon the choice of the positions, a in A and b in B, through which the conjugation is effected. In an earlier paper (Brown 1949), following some work by Coulson and Longuet-Higgins (1948), it was concluded that

$$C_{ab} = -(0.099 + 0.687\sqrt{\lambda_a\lambda_b})\beta, \quad \dots\dots\dots (1)$$

where λ_a and λ_b are the magnitudes of the self-atom-polarizabilities (in units of $1/\beta$) of the positions a and b in A and B respectively.

* Wheatstone Physics Department, King's College, Strand, London, W.C.2.

† We consider throughout the energy of π -electrons only, so that energy changes arising from the making and breaking of σ -bonds are here ignored.

Subsequently it was noticed that the plot of C_{ab} against $\sqrt{\lambda_a \lambda_b}$, although linear in very good approximation (Fig. 1), differed appreciably from that corresponding to (1). This was traced to the use of the Coulson and Longuet-Higgins approximation* $C_{ab} = -p_{ab}\beta_{ab}$. C_{ab} and p_{ab} are indeed linearly related† in very good approximation (Fig. 3), the relationship found by a least-squares fit of the data in Table 1 being

$$p_{ab} = 0.057 - 0.825C_{ab}/\beta, \quad \dots\dots\dots (2)$$

with a standard deviation of 0.002 for p_{ab} .

TABLE 1
DATA USED TO DERIVE CONJUGATION ENERGY EQUATIONS

Molecule	p	$C(-\beta)$	$C'(-\gamma)$	$(\lambda_a \lambda_b)^{\frac{1}{2}}$
Diphenyl	0.370	0.383	0.118	0.398
Styrene	0.406	0.424	0.144	0.446
Butadiene	0.447	0.472	0.175	0.500
Hexatriene	0.483	0.516	0.212	0.559
Octatetraene (end)	0.495	0.530	0.225	0.586
Octatetraene (mid)	0.529	0.573	0.263	0.626
Decapentaene (end)	0.500	0.536	0.231	—
Dodecahexaene (mid)	0.561	0.617	0.294	0.685
Infinite polyene (end)	0.510	0.546	0.242	—
Stilbene	0.431	0.454	0.169	—
1-Phenylbutadiene	—	0.469	0.181	0.499
2-Phenylbutadiene	—	0.385	0.121	0.400
Decapentaene (mid)	—	0.593	0.282	0.655

The fit of the points in Figure 1 to a straight line gives

$$C_{ab} = -(0.811\sqrt{\lambda_a \lambda_b} + 0.062)\beta, \quad \dots\dots\dots (3)$$

with a standard deviation for C_{ab} of 0.003%. This relationship differs appreciably from that previously given (Brown 1949). In Table 2 are given resonance energies, calculated by means of (3), for a number of molecules for some of which the corresponding values have already been derived (Brown 1949) using the less accurate relationship. The differences for these latter, though noticeable, would not appreciably affect qualitative applications to the localization theory of organic reactions, although they might have a very slight influence upon calculations of relative reaction rate constants (Brown 1950a) or equilibrium constants (Brown 1951).

* This is valid when the resonance integral of the linking bond, β_{ab} , is sufficiently small. In the present application we require the conjugation energy for the case where β_{ab} has the standard value for a benzene bond, denoted by β . This value for β_{ab} is sufficiently great to decrease appreciably the accuracy of the Coulson-Longuet-Higgins relationship.

† This is a necessary consequence of the linear relation of Figure 1 and the linear relationship of p_{ab} to $\sqrt{\lambda_a \lambda_b}$ noticed by Coulson and Longuet-Higgins (1948).

In semi-quantitative applications of the localization theory the inclusion of the overlap integral between neighbouring atomic orbitals in the molecular-orbital calculations appears to improve their reliability (Brown 1951). It is therefore important to have a relationship, analogous to (3), for calculating C' .*

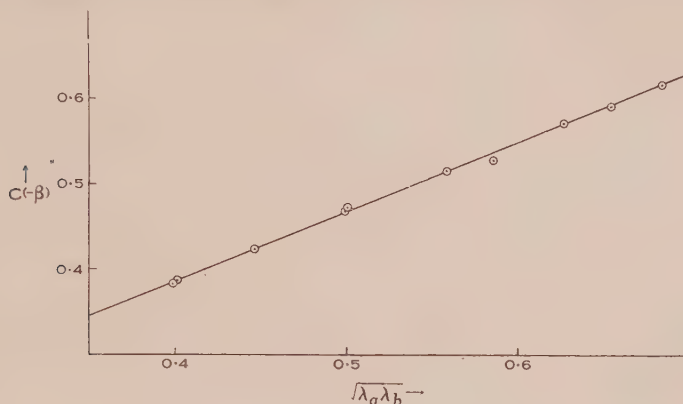


Fig. 1

De Heer (1950), who investigated the calculation of self-atom-polarizabilities when $S=0.25$, showed graphically that C'_{ab} is related approximately linearly to the geometric mean of their magnitudes, λ'_a and λ'_b , but the defect of this relation is that the λ' are more troublesome to compute than the corresponding λ ,

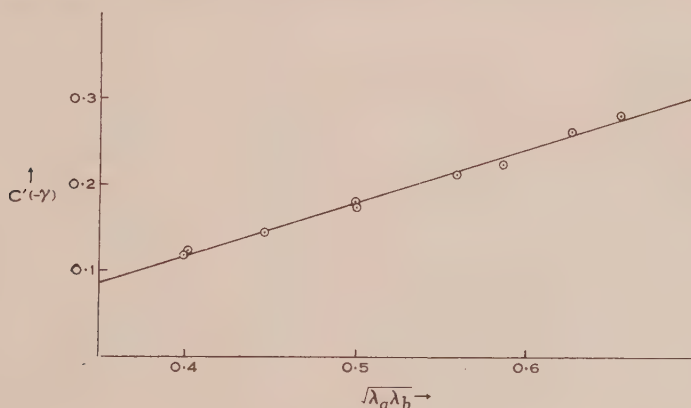


Fig. 2

and very few values of the former are therefore to be found in the literature. However, it was found (Fig. 2) that C'_{ab} is also linearly related to $\sqrt{\lambda_a \lambda_b}$, i.e. the self-polarizability magnitudes *neglecting* overlap. The line of best fit is

$$C'_{ab} = -(0.623\sqrt{\lambda_a \lambda_b} - 0.132)\gamma, \quad \dots\dots\dots (4)$$

the standard deviation for C' being in this case 0.004 γ .

* The use of the prime to denote quantities calculated with the assumption of an overlap integral $S=0.25$ conforms with previous usage (Brown 1949, 1950d, 1952).

For purposes of calculation it seems convenient to tabulate

$$\mu=0.901\sqrt{\lambda}; \quad \mu'=0.789\sqrt{\lambda},$$

so that (3) and (4) may be written

$$C_{ab} = -(\mu_a\mu_b + 0.062)\beta, \quad \dots\dots\dots (5)$$

$$C'_{ab} = -(\mu'_a\mu'_b - 0.132)\gamma, \quad \dots\dots\dots (6)$$

TABLE 2

RESONANCE ENERGIES DERIVED FROM CONJUGATION ENERGIES

Molecule		$C(-\beta)$	$C'(-\gamma)$	$R(-\beta)$	$R'(-\gamma)$
α -Vinyl naphthalene	..	0.444	0.161	4.127	2.024
β -Vinyl naphthalene	..	0.427	0.148	4.110	2.011
α -Phenyl naphthalene	..	0.403	0.130	6.086	3.059
β -Phenyl naphthalene	..	0.388	0.118	6.071	3.048
1-Vinyl anthracene	..	0.448	0.165	5.762	2.773
2-Vinyl anthracene	..	0.430	0.150	5.743	2.758
9-Vinyl anthracene	..	0.478	0.188	5.792	2.795
1-Phenyl anthracene	..	0.407	0.133	7.720	3.807
2-Phenyl anthracene	..	0.390	0.112	7.704	3.786
9-Phenyl anthracene	..	0.433	0.153	7.747	3.828
1-Vinyl phenanthrene	..	0.442	0.160	5.890	2.896
2-Vinyl phenanthrene	..	0.426	0.148	5.874	2.884
3-Vinyl phenanthrene	..	0.430	0.150	5.878	2.886
4-Vinyl phenanthrene	..	0.438	0.157	5.886	2.892
9-Vinyl phenanthrene	..	0.443	0.161	5.892	2.897
1-Phenyl phenanthrene	..	0.401	0.129	7.849	3.932
2-Phenyl phenanthrene	..	0.387	0.118	7.835	3.921
3-Phenyl phenanthrene	..	0.390	0.120	7.838	3.923
4-Phenyl phenanthrene	..	0.397	0.126	7.845	3.928
9-Phenyl phenanthrene	..	0.402	0.130	7.850	3.932
1-Phenyl hexatriene	..	0.485	0.193	3.473	1.647
2-Vinyl hexatriene	..	0.420	0.143	1.408	0.529
2-(β -Styryl)-butadiene	..	0.460	0.174	3.356	1.559

and numerical values of μ and μ' for all positions in various conjugated systems for which the relevant self polarizability data are available in the literature are listed in Table 3.

The calculation of conjugation energies from (5) and (6) may be extended to larger molecules by means of the following approximation. The value of μ (or μ') for the 1-naphthyl position is seen (Table 3) to be very similar to those for 1-anthryl, 1- and 4-phenanthryl; likewise the value for the 2-naphthyl position is very close to those for 2-anthryl, 2- and 3-phenanthryl. Respective values are even closer if corresponding positions in benzanthracene and

TABLE 3
CONJUGATION ENERGY PARAMETERS

Radical	μ	μ'
Vinyl.. ..	0.637	0.558
1-Butadienyl	0.713	0.625
2-Butadienyl	0.571	0.500
1-Hexatrienyl	0.745	0.653
2-Hexatrienyl	0.562	0.492
3-Hexatrienyl	0.618	0.541
1-Octatetraenyl	0.763	0.669
1-Long polyene	0.830	0.728
Middle long polyene	0.637	0.558
Phenyl	0.568	0.498
1-Naphthyl	0.599	0.525
2-Naphthyl	0.573	0.502
1-Anthryl	0.607	0.532
2-Anthryl	0.577	0.506
9-Anthryl	0.653	0.573
1-Phenanthryl	0.597	0.523
2-Phenanthryl	0.572	0.501
3-Phenanthryl	0.577	0.506
4-Phenanthryl	0.590	0.517
9-Phenanthryl	0.599	0.525
2-Diphenyllyl	0.586	0.513
3-Diphenyllyl	0.567	0.497
4-Diphenyllyl	0.578	0.506
1-Diphenylenyl	0.583	0.511
2-Diphenylenyl	0.600	0.525
1'-Benzanthryl	0.597	0.523
2'-Benzanthryl	0.577	0.505
3'-Benzanthryl	0.572	0.501
4'-Benzanthryl	0.597	0.522
3-Benzanthryl	0.603	0.528
4-Benzanthryl	0.603	0.528
5-Benzanthryl	0.605	0.530
6-Benzanthryl	0.576	0.504
7-Benzanthryl	0.577	0.505
8-Benzanthryl	0.604	0.529
9-Benzanthryl	0.634	0.555
10-Benzanthryl	0.646	0.566

anthracene, or benzanthracene and phenanthrene, are compared. Clearly then the fusion of a benzene ring to one end of such aromatic systems has a very small effect upon the self-polarizabilities for positions at the other end of the molecule, the effect rapidly becoming smaller the larger the original system. Thus it is possible to use the known polarizabilities of the smaller molecules in place of the (generally unknown) polarizabilities of larger molecules, e.g. we may use the polarizabilities of the 1- and 2-positions of anthracene for the polarizabilities of the 1- and 2-positions in naphthacene, pentacene, etc. without

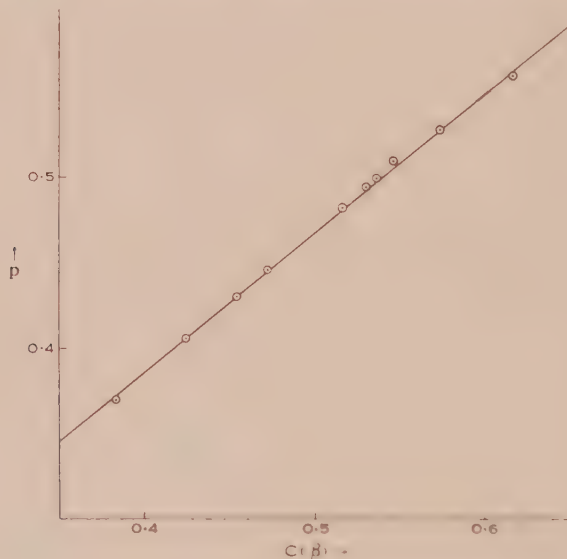


Fig. 3

sensible error for computing conjugation energies. This technique is of particular importance for the calculation of bond localization energies in very large aromatic molecules.

III. EXAMPLE*

To illustrate the application of this technique we shall calculate the bond localization energy of the most reactive bond (i.e. the 3,4-bond) of 1,2,8,9-dibenzopentacene. The residual molecule† corresponding to this bond is 8-phenyl-1,2 benzonaphthacene, so we may compute its resonance energy from the known energies of benzonaphthacene and benzene, together with the conjugation energy of the phenyl group with the 8 position in benzonaphthacene. From equations (5) and (6), with the "large molecule" assumption that the value of λ for the 8-position is sensibly the same as that for the corresponding position (position 6) in benzanthracene, we find

$$C = -0.39\beta; \quad C' = -0.12\gamma,$$

* Numerical data (bond orders, resonance energies, etc.) used in this section are to be found in the literature (Berthier *et al.* 1948; Coulson and Longuet-Higgins 1947; Brown 1950e).

† The terminology and symbolism employed here are described fully by Brown (1949, 1952).

so that, for the residual molecule,

$$R = -11.11\beta; \quad R' = -5.42\gamma.*$$

The resonance energy of the original molecule, dibenzopentacene, is easily calculated from the annelation energy equations (Brown 1950e)

$$R = -12.13\beta; \quad R' = -5.86\gamma,$$

so that, finally, the bond localization energy for the 3 : 4 bond is

$$B = -1.02\beta; \quad B' = -0.44\gamma.$$

Extensive use has already been made of equations (5) and (6), supplemented when necessary by the "large molecule approximation" just described, in the localization treatment of the Diels-Alder and osmium tetroxide reactions of polycyclic hydrocarbons.

IV. ALTERNATIVE APPROXIMATIONS TO CONJUGATION ENERGIES

It seems possible that equation (2), and the analogous equation for C' , may find considerable application for the calculation of conjugation energies, since Coulson and Jacobs (1949) have computed p_{ab} for many linking bonds in such systems, and especially since these authors have devised a particularly convenient technique for such computations. At present, however, equations (5) and (6) seem to cover all cases of interest in the study of chemical reactivities.

V. ACKNOWLEDGMENTS

The author is grateful to Dr. H. H. Greenwood for communicating the self-polarizability data for benzanthracene, incorporated in Table 3, prior to their publication; and to Professor C. A. Coulson, F.R.S., for discussion of the manuscript.

VI. REFERENCES

- BERTHIER, G., COULSON, C. A., GREENWOOD, H. H., and PULLMAN, A. (1948).—*C.R. Acad. Sci. Paris* **226** : 1906.
 BROWN, R. D. (1949).—*Aust. J. Sci. Res.*, A **2** : 564.
 BROWN, R. D. (1950a).—*J. Chem. Soc.* **1950** : 3249.
 BROWN, R. D. (1950b).—*J. Chem. Soc.* **1950** : 691.
 BROWN, R. D. (1950c).—*J. Chem. Soc.* **1950** : 2730.
 BROWN, R. D. (1950d).—*Trans. Faraday Soc.* **46** : 146.
 BROWN, R. D. (1950e).—*Trans. Faraday Soc.* **46** : 1013.
 BROWN, R. D. (1951).—*J. Chem. Soc.* **1951** : 1612.
 BROWN, R. D. (1952).—*Quart. Rev.* **6** : 63.
 COULSON, C. A., and JACOBS, J. (1949).—*J. Chem. Soc.* **1949** : 2805.
 COULSON, C. A., and LONGUET-HIGGINS, H. C. (1947).—*Rev. Sci. Paris* **85** : 929.
 COULSON, C. A., and LONGUET-HIGGINS, H. C. (1948).—*Proc. Roy. Soc. A* **195** : 188.
 DE HEER, J. (1950).—*Phil. Mag.* **41** : 370.

* These compare favourably with the alternative approximate values of -11.10β and -5.42γ obtained from the annelation energy equations (Brown 1950e) for benzanthracene and diphenyl.

PREDICTION CURVES FOR COUNTER-CURRENT SEPARATIONS: APPLICATION TO LIGNIN ALDEHYDES

By D. E. BLAND,* W. E. HILLIS,* and E. J. WILLIAMS†

[Manuscript received December 24, 1951]

Summary

A theoretical investigation of the separation of pairs of solutes in a counter-current distribution train has been made. The results are presented as families of curves from which may be read the number of transfers necessary to effect a 95 or a 99 per cent. separation of a pair of substances of known partition coefficients. The applicability of the curves has been verified when used to predict the number of transfers necessary for the separation of aromatic aldehydes from the oxidative decomposition of lignins. No relation between R_F value and partition coefficient could be established for the solvents investigated.

I. INTRODUCTION

The separation of the aldehydes from the oxidative degradation of lignin by means of nitrobenzene in alkaline medium has hitherto been a laborious process. Creighton, McCarthy, and Hibbert (1941) achieved it by vacuum sublimation. Bland (1949) showed that vanillin and syringaldehyde could be separated by paper chromatography and the method was made quantitative by Stone and Blundell (1951). The aldehydes were also separated on a column of "Magnesol" by Pearl and Dickey (1951). Bland, Ho, and Cohen (1950) used a chromatopile for the isolation of a small quantity of *p*-hydroxybenzaldehyde from the mixture of aldehydes.

The counter-current distribution technique of Craig and Craig (1950, p. 259) appeared to offer an elegant method of separation as well as being applicable on any scale. In the present work the technique was applied to the separation of pairs of the three aldehydes known to be present in the crude aldehydes obtained from lignin.

Johnson (1950) established the relationships governing the transfer of a solute along a train in the case of a solute distributed between immiscible solvents in a series of vessels through which one only of the liquids moves continuously.

Nichols (1950) developed equations for estimating the number of transfers necessary for a given degree of separation. Bacher (1951) presented a formula using the proportions in each tube for estimating the partition coefficient of a solute as a function of counter-current distribution data.

In the present work, it is shown how the numbers of tubes required to effect the separation, to a predetermined degree of accuracy, of two solutes of known

* Division of Forest Products, C.S.I.R.O., Melbourne.

† Section of Mathematical Statistics, C.S.I.R.O., at Division of Forest Products, C.S.I.R.O., Melbourne.

partition coefficients, can be derived from standard tables of the F -distribution used in statistical tests. Curves derived from these tables are presented. Methods of estimating partition coefficients, and the errors to which the estimates are subject, are also discussed.

II. THEORY

In the following theoretical discussion, the operation of a counter-current train is studied under somewhat simplified conditions. No allowance is made for such factors as association of either solute in the different phases. Some factors likely to affect the results are, however, discussed in Section II (*d*). Despite the simplified treatment, the conclusions drawn from the theory give good general agreement with the observed results.

(a) *Estimation of the Number of Transfers Required to Separate Two Solutes*

In any counter-current distribution apparatus, consider a single solute, the fraction of which in the mobile phase in any tube is p . Then p may be regarded as the probability that any particle of the solute will be transferred at a particular transfer. It is a function of the partition coefficient, α , and of the ratio of the volumes of the mobile and stationary phases, V_m and V_s . In fact, if

$$\beta = V_s/V_m,$$

then

$$p = \frac{1}{1 + \alpha\beta}.$$

In the experiments discussed here, $V_m = V_s$, so that

$$p = \frac{1}{1 + \alpha}.$$

After a given number of transfers, the proportions of the solute in the different tubes will have a binomial distribution. If p is known, these proportions are readily calculated. In particular, if N' transfers are carried out, so that there are $N' + 1$ ($= N$) tubes in all, the proportion of solute in the first n_1 tubes is

$$(1-p)^{N'} + N'p(1-p)^{N'-1} + \dots + \binom{N'}{n_1-1} p^{n_1-1} (1-p)^{N'-n_1+1}. \quad (1)$$

Now if two solutes S_1 and S_2 with p -values p_1 and p_2 are in the same solution, they can be satisfactorily separated, provided p_1 and p_2 differ sufficiently. Suppose that p_2 is greater than p_1 . It can be shown that if from N ($= n_1 + n_2$) tubes the first n_1 are retained, and the remaining n_2 discarded, n_1 and n_2 can be so chosen that the amount of S_1 is as large, and the amount of S_2 as small, as desired. This is because the expression (1) decreases with increasing p , its derivative with respect to p being

$$-\frac{p^{n_1-1}(1-p)^{n_2-1}}{B(n_1, n_2)}, \dots \dots \dots (2)$$

where $B(n_1, n_2)$ is the B-function; and this derivative may be made as large numerically as required by choice of n_1 and n_2 .

The most thorough separation of S_1 and S_2 would be achieved by retaining only the first tube for S_1 (and the last tube for S_2). But this method would

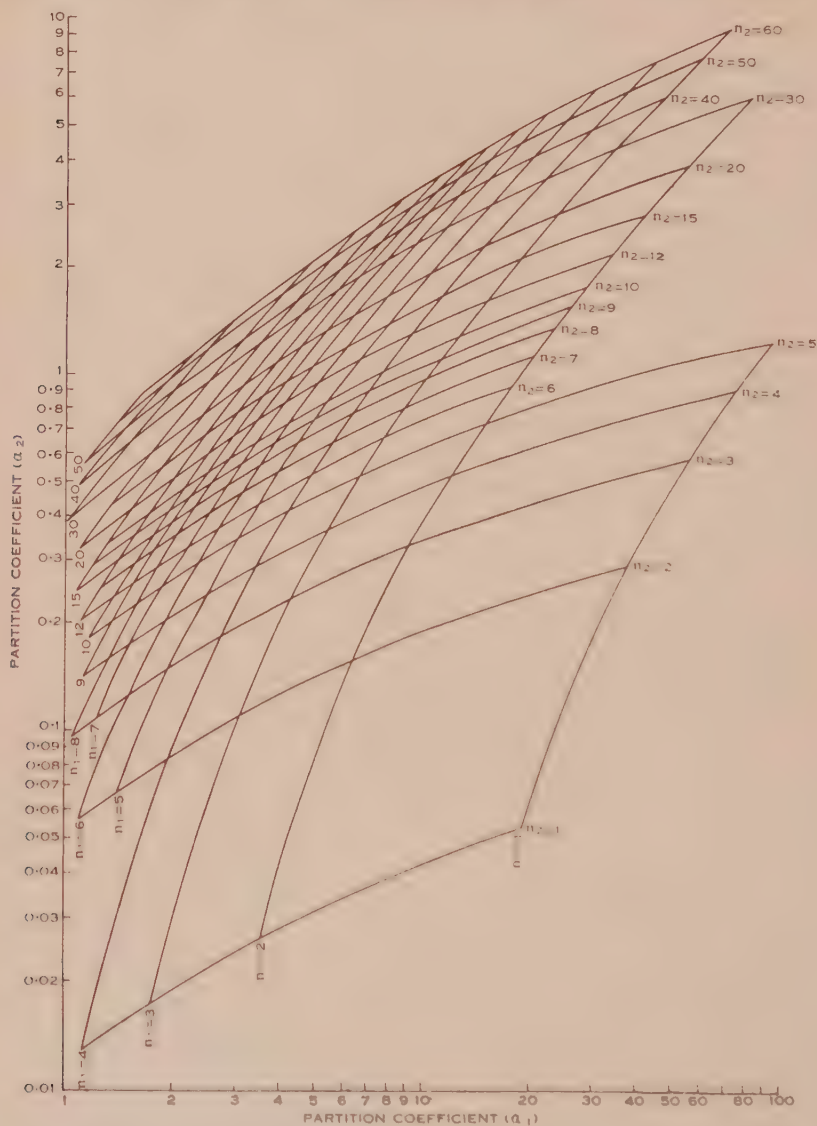


Fig. 1.—Curves showing number of tubes necessary for 99 per cent. separations.

sacrifice much of the material, and would be practicable only if the amount of the mixture available were large. In general, it would be uneconomical. The method which will be considered here is one whereby the first n_1 tubes are retained to give a high proportion of material S_1 , and the remaining n_2 tubes are used to give a high proportion of material S_2 . While not using the full power

of the separation process, this method does conserve all the material, which is important when the material is difficult to prepare.

The characteristics which will be specified for the separation process are, that the first n_1 tubes will contain a fraction $1-u$ of the total amount of material

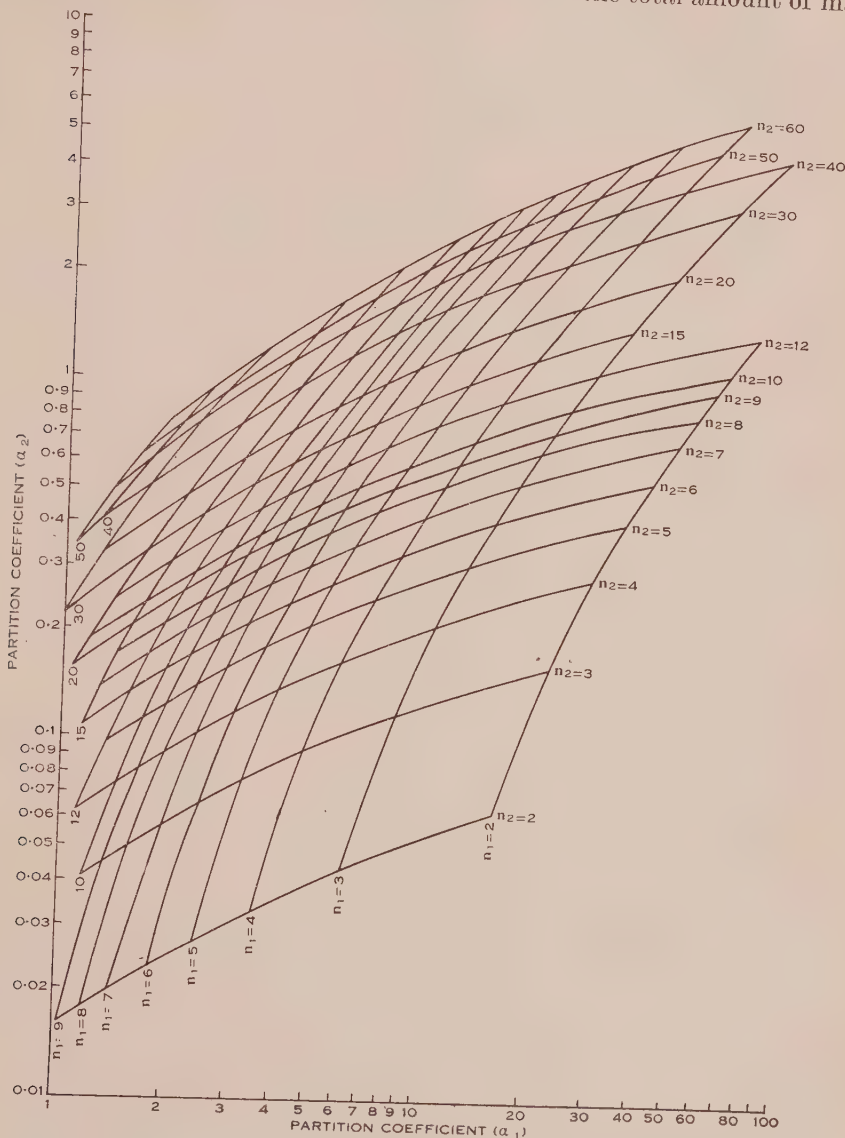


Fig. 2.—Curves showing number of tubes necessary for 95 per cent. separations.

S_1 and that the remaining n_2 tubes will contain a fraction $1-u$ of material S_2 , where u is small. The contamination of each material will then be by a fraction u of the total amount of the other. In practice, u is taken as 0.05 or 0.01.

For given values of n_1 and n_2 , it is possible to determine the values of p_1 and p_2 for which the process will fulfil these conditions. For $p = p_1$, it is required

that the sum (1) shall equal $1-u$. On integrating the expression (2), it is found that (1) may be expressed as the integral

$$\int_p^1 \frac{s^{n_1-1}(1-s)^{n_2-1}}{B(n_1, n_2)} ds, \dots\dots\dots (3)$$

which is the probability that a variate having an F -distribution with $2n_2$ and $2n_1$ degrees of freedom shall be less than

$$\frac{n_1(1-p)}{n_2p}.$$

Such values of F are tabulated by Fisher and Yates (1949, Table 5), for $u=0.20, 0.10, 0.05, 0.01, 0.001$, so that to find p_1 we have simply

$$F_1 = \frac{n_1(1-p_1)}{n_2p_1},$$

$$p_1 = \frac{n_1}{n_2F_1 + n_1},$$

$$\alpha_1 = \frac{n_2F_1}{n_1}.$$

In the same way, it is seen that, for $p=p_2$, the sum (1) must equal u , which is then the probability that a variate having an F -distribution with $2n_1$ and $2n_2$ degrees of freedom shall exceed

$$\frac{n_2p}{n_1(1-p)}.$$

Calling this value F_2 , we have

$$p_2 = \frac{n_1F_2}{n_1F_2 + n_2},$$

$$x_2 = \frac{n_2}{n_1F_2}.$$

As an example, consider what materials could satisfactorily be separated using 3 tubes, with $u=0.05$, and taking

$$n_1=2,$$

$$n_2=1,$$

then

$$F_1=6.944, \text{ (2 and 4 degrees of freedom)}$$

$$x_1 = \frac{6.944}{2} = 3.472.$$

$$F_2=19.247, \text{ (4 and 2 degrees of freedom)}$$

$$x_2 = \frac{1}{2 \times 19.247} = 0.0260.$$

These results could also be obtained directly. For, proportion of material in first two tubes $= 1 - p^2$, hence for material S_1

$$\begin{aligned} 1 - p_1^2 &= 0.95, \\ p_1^2 &= 0.05, \\ p_1 &= 0.2236, \\ \alpha_1 &= \frac{0.7764}{0.2236} = 3.472; \end{aligned}$$

while, for material S_2

$$\begin{aligned} p_2^2 &= 0.95, \\ p_2 &= 0.9747, \\ \alpha_2 &= \frac{0.0253}{0.9747} = 0.0260. \end{aligned}$$

As is to be expected, only materials whose α -values differ greatly can be satisfactorily separated with so small a number of tubes.

Figures 1 and 2 show the values of n_1 and n_2 required to separate two materials, given the values α_1 and α_2 ; Figure 1 for $u = 0.05$, and Figure 2 for $u = 0.01$. From these, the number of transfers required in a given separation can be readily determined.

(b) Approximations When N is Large

Corresponding to most values of n_1 and n_2 in the practical range values of F can be read from published tables. For values of n_1 and n_2 outside the range of the tables Fisher has given approximate formulae for determining F . These are as follows:

If $z = \frac{1}{2} \log_e F$, we have

$$\begin{aligned} u = 0.05 \quad & \left| \begin{aligned} z_1 &= \frac{1.6449}{\sqrt{h-1}} + 0.7843 \left(\frac{1}{2n_1} - \frac{1}{2n_2} \right) \\ z_2 &= \frac{1.6449}{\sqrt{h-1}} - 0.7843 \left(\frac{1}{2n_1} - \frac{1}{2n_2} \right) \end{aligned} \right| \\ u = 0.01 \quad & \left| \begin{aligned} z_1 &= \frac{2.3263}{\sqrt{h-1.4}} + 1.2353 \left(\frac{1}{2n_1} - \frac{1}{2n_2} \right) \\ z_2 &= \frac{2.3263}{\sqrt{h-1.4}} - 1.2353 \left(\frac{1}{2n_1} - \frac{1}{2n_2} \right) \end{aligned} \right| \end{aligned}$$

where

$$h = \frac{4n_1n_2}{n_1 + n_2}.$$

$\log_{10} F$ is obtained by multiplying these values by 0.8686, whence F can be read from a table of logarithms.

The above approximation has the disadvantage in this application that it gives α_1 and α_2 in terms of n_1 and n_2 , whereas generally n_1 and n_2 are required in terms of α_1 and α_2 . Formulae of this latter type will therefore be developed using a different approach.

Consider a variate x having a binomial distribution, with parameters p and N' . Then x ranges from 0 to N' with mean $N'p$ and variance $N'p(1-p)$.

It will be convenient to number the tubes in the counter-current distribution 0, 1, 2, . . . N' , to correspond to the values of x . Then for a solute giving a fraction p in the mobile phase, the proportion in the r th tube is equal to the probability that $x=r$.

If now we find the values of x cutting off a fraction u from either tail of the distribution, this will give the number of tubes containing, in one case a fraction $1-u$, and in the other, a fraction u of the solute. To do this we use the method of asymptotic expansion (see, for example, Kendall 1948, p. 158). Let

$$v^2 = p(1-p),$$

and

$$y = \frac{x - N'p}{N'^{\frac{1}{2}}v}.$$

Then y is a variate with zero mean and unit standard deviation: its third cumulant is

$$\frac{1-2p}{N'^{\frac{1}{2}}v}.$$

If ξ is the normal deviate corresponding to the rejection probability u , then the required value of y is given by the asymptotic formula

$$y = \xi + \frac{1-2p}{6N'^{\frac{1}{2}}v}(\xi^2-1) + o(N'^{-\frac{1}{2}}), \quad \dots\dots\dots (4)$$

whence

$$x = N'p + N'^{\frac{1}{2}}v\xi + \frac{1-2p}{6}(\xi^2-1) + o(1). \quad \dots\dots\dots (5)$$

The normal deviate corresponding to the rejection probability $1-u$ is $-\xi$; hence, changing the sign of ξ in (5) gives the lower probability point of the distribution.

To separate two given solutes S_1 and S_2 , with fractions p_1 and p_2 , it is necessary so to choose N' that the upper probability point for S_1 corresponds with the lower probability point for S_2 . The common probability point then divides the distribution into the required n_1 and n_2 tubes.

We put

$$p_1 = \sin^2 \theta_1, \quad p_2 = \sin^2 \theta_2;$$

so that

$$\alpha_1 = \cot^2 \theta_1, \quad \alpha_2 = \cot^2 \theta_2.$$

Then

$$\begin{aligned} p_2 - p_1 &= \sin^2 \theta_2 - \sin^2 \theta_1, \\ &= \sin (\theta_2 + \theta_1) \sin (\theta_2 - \theta_1). \quad \dots\dots\dots (6) \end{aligned}$$

Also

$$\begin{aligned} v_2 + v_1 &= \sin \theta_2 \cos \theta_2 + \sin \theta_1 \cos \theta_1, \\ &= \sin (\theta_2 + \theta_1) \cos (\theta_2 - \theta_1). \quad \dots\dots\dots (7) \end{aligned}$$

(i) *First Approximation.*—In (5), we ignore terms $O(1)$.

$$N'p_1 + N'^{\frac{1}{2}}v_1\xi = N'p_2 - N'^{\frac{1}{2}}v_2\xi, \quad \dots\dots\dots (8)$$

so that

$$N' = \xi^2 \left(\frac{v_2 + v_1}{p_2 - p_1} \right)^2,$$

from (6) and (7)

$$= \xi^2 \cot^2 (\theta_2 - \theta_1). \quad \dots \dots \dots (9)$$

(ii) *Second Approximation.*—We now include terms $O(1)$; then

$$N'p_1 + N'^{\frac{1}{2}}v_1\xi + \frac{(1-2p_1)}{6}(\xi^2-1) = N'p_2 - N'^{\frac{1}{2}}v_2\xi + \frac{(1-2p_2)}{6}(\xi^2-1), \dots (10)$$

which reduces to

$$N'(p_2 - p_1) - N'^{\frac{1}{2}}\xi(v_2 + v_1) - \frac{(p_2 - p_1)}{3}(\xi^2 - 1) = 0,$$

or

$$N' - N'^{\frac{1}{2}}\xi \cot (\theta_2 - \theta_1) - \frac{\xi^2 - 1}{3} = 0; \quad \dots \dots \dots (11)$$

and the relevant solution is

$$N'^{\frac{1}{2}} = \frac{1}{2} \left[\xi \cot (\theta_2 - \theta_1) + \sqrt{\xi^2 \cot^2 (\theta_2 - \theta_1) + \frac{4}{3}(\xi^2 - 1)} \right], \quad \dots (12)$$

that is,

$$N' = \frac{1}{2} \left[\xi^2 \cot^2 (\theta_2 - \theta_1) + \frac{2}{3}(\xi^2 - 1) + \xi \cot (\theta_2 - \theta_1) \sqrt{\xi^2 \cot^2 (\theta_2 - \theta_1) + \frac{4}{3}(\xi^2 - 1)} \right], \quad \dots \dots \dots (13)$$

$$= \xi^2 \cot^2 (\theta_2 - \theta_1) + \frac{2}{3}(\xi^2 - 1) + O(N'^{-1}). \quad \dots \dots \dots (13')$$

Further terms of (5) may be included for greater accuracy, but approximation given by (13) is good enough in most cases. It is seen from (13') that the error of equation (9) is about $\frac{2}{3}(\xi^2 - 1)$, that is between 1 and 2 tubes. The value of $n_1 - \frac{1}{2}$ is now given by the left-hand side of (10). That is, on substituting for N' from (11)

$$n_1 = \frac{1}{2} + \sin \theta_1 \sin \theta_2 \operatorname{cosec} (\theta_2 - \theta_1) N'^{\frac{1}{2}} \xi + \frac{\xi^2 - 1}{6}, \quad \dots (14)$$

$$n_2 = \frac{1}{2} + \cos \theta_1 \cos \theta_2 \operatorname{cosec} (\theta_2 - \theta_1) N'^{\frac{1}{2}} \xi + \frac{\xi^2 - 1}{6}, \quad \dots (15)$$

where $N'^{\frac{1}{2}}$ is given by (11). Hence

$$\frac{n_2 - (\xi^2 + 2)/6}{n_1 - (\xi^2 + 2)/6} = \cot \theta_2 \cot \theta_1 = \sqrt{\alpha_2 \alpha_1}, \quad \dots \dots \dots (16)$$

and n_1, n_2 are generally more easily found from this equation.

For $u = 0.05$, $\xi = 1.6449$, and we have

$$\frac{n_2 - 0.7843}{n_1 - 0.7843} = \cot \theta_2 \cot \theta_1 = \sqrt{\alpha_2 \alpha_1}.$$

For $u=0.01$, $\xi=2.3263$, and we have

$$\frac{n_2-1.2353}{n_1-1.2353} = \cot \theta_2 \cot \theta_1 = \sqrt{z_2 z_1}.$$

We note in passing that, when $\alpha_2=1/\alpha_1$, $p_1+p_2=1$, and $n_1=n_2$.

Examples

(i) Suppose

$$z_1=9.00, \quad z_2=0.25,$$

then

$$\cot (\theta_2-\theta_1)=1.$$

From (13)

$$\begin{aligned} N' &= \frac{1}{2} \left[\xi^2 + \frac{2}{3}(\xi^2-1) + \xi \sqrt{\xi^2 + \frac{4}{3}(\xi^2-1)} \right] \\ &= 4(3.76) \quad \text{for } u=0.05, \\ &= 8(8.09) \quad \text{for } u=0.01. \end{aligned}$$

The values of n_1 and n_2 , from Figure 1, are 2 and 3, giving $N'=4$, and from Figure 2, are 4 and 5, giving $N'=8$.

It is seen that equation (13) gives an adequate approximation, even in this range.

(ii) Suppose

$$z_1=2.25, \quad z_2=1.00,$$

then

$$\cot (\theta_2-\theta_1)=5.$$

From (13)

$$\begin{aligned} N' &= \frac{1}{2} \left[25\xi^2 + \frac{2}{3}(\xi^2-1) + 5\xi \sqrt{25\xi^2 + \frac{4}{3}(\xi^2-1)} \right] \\ &= 69 \quad \text{for } u=0.05, \\ &= 138 \quad \text{for } u=0.01. \end{aligned}$$

From (15) for

$$u=0.05,$$

$$n_1=28,$$

$$n_2=12;$$

for

$$u=0.01,$$

$$n_1=56,$$

$$n_2=83.$$

(c) Separation of Solutes using a Fixed Number of Tubes

Where the partition coefficients of a pair of solutes do not differ sufficiently, the number of transfers required to separate them to a given accuracy by the methods outlined in the previous sections may be prohibitively large. In such cases, it may be necessary to carry out a number of transfers, and to reject the middle tubes in which the contamination is greatest. The reduction in time

and apparatus required compensates for the loss of material involved; furthermore, if the material from the middle tubes is sufficient to warrant further treatment, it may be put through a subsequent separation process.

For separation with rejection of middle tubes, N , the number of tubes to be used, is first decided. Then it is specified that the first v_1 tubes be contaminated with a fraction u of the total amount of solute S_2 , and that the last v_2 tubes be contaminated with a fraction u of the total amount of solute S_1 . The remaining $v_0 (=N-v_1-v_2)$ tubes are to be rejected.

This specification is not as stringent as that given in Section II (a) since, while the first v_1 tubes contain a fraction u of the contaminant, they contain less than a fraction $1-u$ of the solute S_1 . It can be tightened up if necessary by the use of a smaller value of u .

To use Figure 1 or 2 with this procedure, draw on the chart the curve $n_1 - n_2 = N$ (constant). Where this curve cuts the abscissa for α_2 , the value of n_1 gives v_1 , and where it cuts the ordinate for α_1 , the value of n_2 gives v_2 .

Alternatively, the method of asymptotic expansion may be used. From the results of the Section II (b) (ii), with N fixed, it is seen that the right-hand side of (10) gives $v_1 - \frac{1}{2}$:

$$v_1 = \frac{1}{2} - N'p_2 - N'^{\frac{1}{2}}v_2\zeta + \frac{(1-2p_2)}{6}(\zeta^2-1); \quad \dots\dots\dots (17)$$

similarly,

$$v_2 = \frac{1}{2} + N'(1-p_1) - N'^{\frac{1}{2}}v_1\zeta - \frac{(1-2p_1)}{6}(\zeta^2-1). \quad \dots\dots (18)$$

The number of tubes rejected is

$$\begin{aligned} v_0 &= N' - v_1 - v_2 + 1 \\ &= -N'(p_2 - p_1) + N'^{\frac{1}{2}}\zeta(v_2 + v_1) - \frac{(p_2 - p_1)}{3}(\zeta^2 - 1), \quad \dots (19) \end{aligned}$$

which is zero, as it should be, when N' is a solution of (11).

Example

Let $N=20$, $\alpha_1=2$, $\alpha_2=1$, and $u=0.01$. From Figure 2

$$v_1=5,$$

$$v_2=8,$$

so that

$$v_0=7.$$

From (16) and (17) since $p_1=1/3$ and $p_2=1/2$

$$v_1 = \frac{1}{2} + \frac{19}{2} - \frac{\sqrt{19}}{2} \times 2.3263 + 0,$$

$$= 5(4.93);$$

$$v_2 = \frac{1}{2} + \frac{38}{3} - \frac{\sqrt{38}}{3} \times 2.3263 - \frac{4.4119}{18},$$

$$= 8(8.14).$$

The formulae give good agreement with the Figures 1 and 2. The method is not applied to the results discussed here, but the example shows how it could be used.

(d) *Effect of Errors in Estimation of the Partition Coefficient
and in the Operation of the Apparatus*

The partition coefficient for any solute in any pair of solvents is seldom known accurately. Moreover, it is often determined from results of experiments other than the counter-current distribution. Each form of experiment is subject to its own characteristic errors. For instance, in counter-current distribution work it is found that portion of the stationary phase tends to be carried over from one tube to the next with the mobile phase. The quantity of solute in the first tube is therefore lower than that theoretically expected. As a result of such disturbances, the value of the partition coefficient determined from one type of experiment does not apply exactly to another. For such reasons it is desirable that the partition coefficient of a substance for counter-current distribution work should be determined from a counter-current distribution experiment. Methods of estimating the partition coefficient are discussed below; in this section, the effect of errors in the estimated coefficient on the proportions obtained in the separation process is considered.

Suppose that the estimated value of a partition coefficient α is α' ; the error in the partition coefficient is $h = \alpha' - \alpha$, and that in the fraction p ,

$$\frac{\alpha' - \alpha}{(1 + \alpha)(1 + \alpha')} \sim \frac{h}{(1 + \alpha)^2}$$

Now since the proportion of solute in the first n_1 of N tubes is

$$\int_p^1 \frac{t^{n_1-1}(1-t)^{n_2-1}}{B(n_1, n_2)} dt,$$

an error h in α introduces an error

$$\frac{h}{(1 + \alpha)^2} \frac{p^{n_1-1}(1-p)^{n_2-1}}{B(n_1, n_2)} = \frac{h\alpha^{n_1-1}}{B(n_1, n_2)(1 + \alpha)^N}$$

in the estimate of this proportion. Hence, an overestimated value of α will lead to an overestimated value of the proportion in the first n_1 tubes (and an underestimated value of the proportion in the remaining n_2 tubes). The contamination will be less than that predicted if α_1 is underestimated, and α_2 is overestimated.

Errors due to the transfer of portion of the stationary phase are now considered. For simplicity, it is assumed that a fraction φ of the stationary phase is transferred with the mobile phase at each transfer. Then the fraction of solute transferred from one tube to the next in one transfer is

$$\begin{aligned}\hat{p} &= p + \varphi(1 - p), \\ &= p + \varphi - \varphi p.\end{aligned}$$

Hence the net result of the transfer is the same as that of increasing the fraction from p to \hat{p} ; correspondingly, the value of α is reduced to

$$\alpha = \frac{z(1-\varphi)}{1-\varphi z} \quad \dots\dots\dots (20)$$

The observed proportions should be found to agree with theory, if z is replaced by $\hat{\alpha}$.

In several of the experiments comprising the work which is described below, discrepancies between observed and theoretical proportions have been found, even when the values of z have been estimated from the data so as to give as good agreement as possible. It is possible that the discrepancies are due partly to variations in the value of z at different concentrations. (See for instance Craig and Craig 1950, p. 174). Other discrepancies may be attributed to errors in the experimental and analytical methods. To the estimation of partition coefficients, and the assessment of the errors of the method, we must now turn.

(e) *Estimation of Partition Coefficients from Counter-Current Distribution Results*

As mentioned above, it is desirable to determine partition coefficients for this work from actual counter-current distribution trials, so that any factors in the technique which influence the estimate will also be present in the subsequent experiments.

Suppose that, in a trial run, N tubes are used, and that n tubes contain measurable quantities of the solute being studied. Let the quantities, expressed as a fraction of the total solute, be

$$w_0, w_1, w_2, \dots, w_{N'},$$

so that

$$\sum_{r=0}^{N'} w_r = 1.$$

Then the expected value of w_r is

$$P_r = \binom{N'}{r} p^r (1-p)^{N'-r} \quad \dots\dots\dots (21)$$

In general, w_r will vary about P_r owing to errors of sampling and technique; the observed values, however, provide estimates of the P_r , and hence of p .

A rough estimate of p is given by the location of the maximum value of w_r . If this occurs in tube m , the estimate is

$$p_m = \frac{N' - 1}{m - \frac{1}{2}} \quad \dots\dots\dots (22)$$

This formula is slightly more accurate than that given by Craig and Craig (1950, p. 207), which, in our notation, leads to the estimate m/N' for p .

An estimate which is more satisfactory, in that it makes use of information provided by all the tubes, is

$$p' = \Sigma r w_r / N' \quad \dots\dots\dots (23)$$

This is the estimate which would be used if the w_r were in fact a sample from a binomial distribution. In the present application, the distribution is developed by the random transfer of a large number of small particles, and the binomial variation is therefore negligible. Each value of w_r has a distribution which is virtually continuous, the important variations in which are the result of random errors in the experimental method. The form and magnitude of these variations cannot be stated *a priori*, and have to be determined from the data. This point appears to have been overlooked by Bacher (1951), who describes his equation (1), equivalent to our formula (23), as being the maximum likelihood estimate of p . The method of maximum likelihood leads to this formula only if the variation is, in fact, binomial.

However, it will here be assumed, pending further information, that the quantities w_r are distributed according to the following:

$$dF = \text{const.} \prod_{r=0}^{N'} w_r^{\lambda P_r - 1} \prod_{r=0}^{N'-1} dw_r, \quad \dots \dots \dots (24)$$

with

$$\sum_{r=0}^{N'} w_r = 1;$$

here λ is a scale factor, which is to be estimated from the data. Since λ is large in these applications, in which the errors of the observations are small, only the highest power of λ in any expression is retained. Then the variances and covariances of the w_r are

$$V(w_r) = P_r(1 - P_r)/\lambda, \quad \dots \dots \dots (25)$$

and

$$\text{Cov}(w_r, w_s) = -P_r P_s / \lambda. \quad \dots \dots \dots (26)$$

With these assumptions, it can be shown that the estimate p' is efficient: that is, its variance is almost as small as possible. The variance of the estimate p' is

$$V(p') = p(1 - p)/N'\lambda. \quad \dots \dots \dots (27)$$

It now remains to estimate λ . It can be shown that an efficient estimate of λ is given by λ' , where

$$\frac{1}{\lambda'} = \frac{2}{n-1} \sum_{r=0}^{N'} P_r \log_e (P_r/w_r); \quad \dots \dots \dots (28)$$

for practical purposes, where great accuracy is not required, a simpler estimate is given by λ'' , where

$$\frac{1}{\lambda''} = \frac{1}{n-2} \left(\sum_{r=0}^{N'} w_r^2 / p_r - 1 \right). \quad \dots \dots \dots (29)$$

In both these estimates, P_r is to be calculated from p' , the estimate of p provided by the data.

Substitution of either estimate of λ in (27) gives the variance of p' . With the variance of p' known, confidence limits for p can be derived, and hence, limits for x and the proportion of solute in the first n_1 tubes, by the methods described in Section II (d).

(f) *Information about p Provided by Results for Different Tubes*

The determination of the amount of solute in each tube takes considerable time, and it is therefore important to see whether certain determinations can be omitted without reducing too greatly the accuracy of the estimate of p .

The information about an unknown quantity provided by any set of measurements is defined as the inverse of the variance of an efficient estimate. In the estimation of p , the information provided by a distribution in N tubes is seen, from equation (27), to be

$$N'\lambda p(1-p), \quad \dots\dots\dots (30)$$

Hence, roughly speaking, each tube after the first adds $\lambda/p(1-p)$ units of information to the total. However, different tubes in any experiment provide different amounts of information. The information provided by tube r can be shown to be

$$\frac{\lambda(r-N'p)^2 P_r}{p^2(1-p)^2}, \quad \dots\dots\dots (31)$$

which is small for r in the neighbourhood of 0, $N'p$, or N' .

It can be seen by comparison of (22) and (31) that, while the *position* of the maximum value of w_r provides some information about p , the *value* of the maximum provides practically none. This is because, as can readily be seen, variation of w_m has little effect on p' .

The values of r which maximize the information are found to be, to $O(1)$,

$$N'p \pm \sqrt{2N'p(1-p)} + \frac{1-2p}{4}, \quad \dots\dots\dots (32)$$

and the information given by each tube in the neighbourhood of these two values is about

$$\frac{\lambda}{ep(1-p)} \sqrt{\frac{2N'}{\Pi p(1-p)}}, \quad \dots\dots\dots (33)$$

When p is small, $o(N^{-1})$, most of the information is given by tube 1. For then the information from the successive tubes can be written

$$I_0 = N'\lambda (N' - N'(N'-2)p + \dots),$$

$$I_1 = N'\lambda \left(p^{-1} - 3(N'-1) - (N'-1)(7N'-12)\frac{p}{2} - \dots \right),$$

$$I_2 = N'\lambda (2(N'-1) - 4(N'-1)(N'-2)p + \dots),$$

etc., the total information being given by

$$I = N'\lambda (p^{-1} + 1 + p + \dots);$$

and I_1 alone is $O(N'p^{-1})$.

The formulae (31) and (33) actually apply only when a number of tubes is measured. For clearly, measurement of a single tube provides no information at all about p . If two tubes, r and s , are measured, the total information provided is found to be

$$\frac{\lambda P_r P_s (r-s)^2}{P_r + P_s}, \quad \dots\dots\dots (34)$$

When r and s take the two values given by (32), this information becomes

$$\frac{2\lambda}{e} \sqrt{\frac{2N'p(1-p)}{\pi}} \dots\dots\dots (35)$$

If any two observations x_r, x_s are taken, their ratio provides an estimate of $p/(1-p)$;

$$\frac{\binom{n}{s} x_r}{\binom{n}{s} x_s} \sim \left(\frac{p}{1-p} \right)^{r-s} \dots\dots\dots (36)$$

Equating the two sides of (36) leads to an estimate of p , and the reciprocal of (34) gives its variance. When more than two, but fewer than N tubes are measured, the estimation formulae become more complicated, and cannot be given here.

III. PROCEDURE

The applicability of the theoretical development given above was tested in the following manner.

The partition coefficients of vanillin, syringaldehyde, and *p*-hydroxybenzaldehyde between various solvents and water were determined. These values were used to predict from Figures 1 and 2 the number of transfers necessary to effect the separation of pairs of the aldehydes with different solvents by counter-current distribution. Counter-current separation was then carried out with the predicted number of transfers, after which aliquots were withdrawn from the aqueous layer in each tube and analysed for each of the two substances. The percentage of the total amount of each substance in the first n_1 and the last n_2 tubes was calculated for each separation and the results compared with the prediction.

In order to test the curves with reference to a different class of compound a separation of maleic and fumaric acids was carried out with the number of transfers predicted by the curves from published partition coefficients.

An attempt was made to establish a relationship between partition coefficient and thence counter-current distribution results with R_F values. Therefore the R_F values of the three aldehydes in the same solvents were determined as accurately as possible under rigidly standardized conditions.

IV. EXPERIMENTAL

(a) Measurement of Partition Coefficients

The aldehyde being investigated (0.5 g. of vanillin, 0.5 g. of *p*-hydroxybenzaldehyde, or 0.25 g. of syringaldehyde) was distributed between the aqueous and organic phases (100 ml. of each) by mechanically shaking vigorously for several hours. An aliquot (10 ml.) of each phase was measured from a burette and made up to 100 ml. with ethanol. Aliquots (10 ml.) of this solution were analysed by the method of Iddles *et al.* (1939), the precipitate of 2,4-dinitrophenylhydrazone being collected after 24 hours. The washed precipitate was dried to constant weight in a vacuum oven (45°C.).

Under exactly the same conditions, control determinations were done on each of the phases containing a known amount of the aldehyde being investigated (0.2 g. of vanillin, 0.2 g. of *p*-hydroxybenzaldehyde, or 0.1 g. of syringaldehyde). All determinations were done in duplicate at room temperature (c. 25 °C.).

The partition coefficient was calculated as the ratio of the concentration of aldehyde in the aqueous phase to the concentration in organic phase.

(b) *Determination of R_F Values*

"Shell X222 solvent" (b.p. 40–100 °C.) from one drum only was used without any purification and the other components were A.R. benzene and distilled water. The aldehydes (c. 0.5 mg.) were placed 2.5 cm. from the edge of the paper (No. 1 Whatman) which was then clipped into a cylindrical shape by means of staples. The chromatographic solvents were allowed to stand for about 1 hour before the phases were separated. The aqueous phase was poured on to the bottom of a stainless steel tank (12 in. diam. by 21 in.) and a stoppered flask containing the organic phase and a glass dish placed in position. A stainless steel plate was clipped over the mouth of the tank and the protecting lid rested in the water seal on the outside of the tank. The tank was kept in a constant temperature room (22 °C.). After 20 hours the paper was stood in the glass dish and conditioned for 22 hours. The organic solvent was then carefully poured into the dish and allowed to irrigate the paper for 8 hours. The solvent boundary was detected under ultraviolet light and the position of the aldehydes determined by spraying with 2,4-dinitrophenylhydrazine hydrochloride. The procedure was repeated for each individual determination.

(c) *Testing of the Curves*

Four pairs of substances were taken: vanillin and syringaldehyde, vanillin and *p*-hydroxybenzaldehyde, syringaldehyde and *p*-hydroxybenzaldehyde, and fumaric and maleic acids. The number of transfers necessary to give separation was read from the curves in Figures 1 and 2. For example, in the case of syringaldehyde ($\alpha_1=11.3$) and vanillin ($\alpha_2=4.3$), it is indicated in Figure 1* that 95 per cent. of the syringaldehyde and 5 per cent. of vanillin is to be expected in the first 12 tubes (that is, Nos. 0–11) and the remainder of these substances in the last 70 tubes (that is, Nos. 12–81). The separation was then attempted with the aid of a 23-tube Craig machine (tubes taking 20 ml. of each phase of the equilibrated solvent) using about 0.1 g. of each solvent dissolved in the aqueous phase in the first tube. After the estimated number of transfers had been completed, aliquots of the aqueous layer of each tube were analysed for each of the substances. This was done by making the solutions alkaline and measuring the absorption at two wavelengths in the ultraviolet region on a Unicam spectrophotometer. For the determination of vanillin and syringaldehyde in the presence of one another, absorption measurements were taken at 340 and 380 m μ ; for *p*-hydroxybenzaldehyde and syringaldehyde at 320 and 380 m μ ; and for *p*-hydroxybenzaldehyde and vanillin at 320 and

* In this case by a short extrapolation.

360 m μ . After taking measurements on solutions of known concentration of each of the aldehydes at the appropriate wavelengths the amount of each substance in each tube was calculated from the pair of equations so obtained.

V. RESULTS AND DISCUSSION

(a) *Comparison of Results with Predictions*

The partition coefficients measured for the substances in different solvents are shown in Table 1. The predicted number of tubes necessary for separations at the levels indicated in Table 1 were read from Figures 1 and 2. The actual

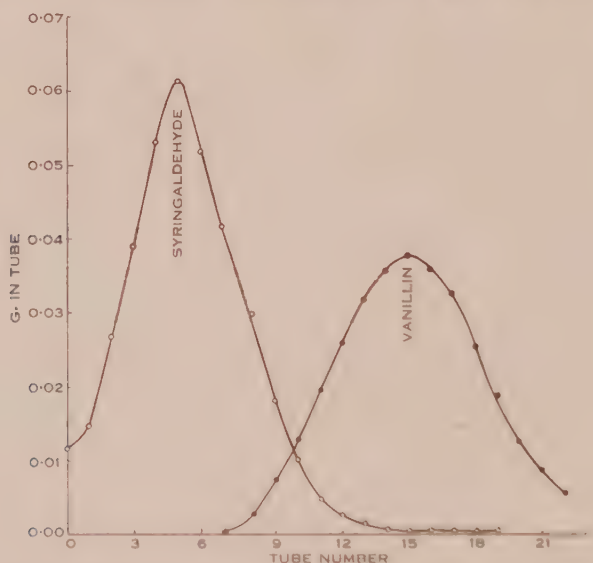


Fig. 3.—Separation of syringaldehyde and vanillin. "Shell X222 solvent":isopropyl ether-water, 82 transfers.

separation achieved is shown in the last two columns as percentages of each substance found in the first n_1 and in the last n_2 tubes. The specimen distribution curve shown in Figure 3 is for vanillin and syringaldehyde separated by the use of "X222":isopropyl ether mixture with 82 transfers. This curve is shown in order to illustrate some points in relation to the predictions, from Figures 1 and 2. It can be seen by reference to Figure 3 that, although a number of transfers equal to $n_1 + n_2$ was necessary to effect the separation, the solute of partition coefficient α_1 was distributed over the first 10 tubes but the solute of partition coefficient α_2 was mainly in tubes 10 to 22, only a negligible amount having passed tube 23. In other words the $n_1 + n_2$ is the number of transfers necessary for separation but no information concerning which tubes will contain significant amounts of the solutes is available from the curves. Where, as in this case, $n_1 + n_2$ is large, the curves can be read only to within several tubes and it is to be expected that the result might not be exactly as predicted. The separation of vanillin from syringaldehyde was in fact close to a 95 per cent. separation as predicted but only when n_1 was taken as 10 instead of 12. This

is shown in Table 1. Reference to this table will also show that separation of other pairs of aldehydes took place as predicted from the curves. The results of the separation of fumaric and maleic acids using ethyl ether as the moving

TABLE 1
RESULTS OF SEPARATIONS WITH PREDICTED NUMBERS OF TUBES

Pairs of Substances	Solvent	Parti- tion Co- efficients	Degree of Separa- tion Sought (%)	Pre- dicted Neces- sary Number of Tubes (n_1+n_2)	Found in First n_1 Tubes (%)	Found in Last n_2 Tubes (%)
<i>p</i> -Hydroxybenzaldehyde Vanillin	Benzene	2.1 0.17	99	10+7	99.4 1.4	0.6 98.6
<i>p</i> -Hydroxybenzaldehyde Syringaldehyde ..	Benzene	2.1 0.14	99	9+6	99.5 2.0	0.5 98.0
<i>p</i> -Hydroxybenzaldehyde Syringaldehyde ..	"X222": Benzene 20:80	4.3 0.38	99	8+10	99.6 0.9	0.4 99.1
<i>p</i> -Hydroxybenzaldehyde Vanillin	"X222": Benzene 20:80	4.3 0.21	99	6+6	99.5 2.5	0.5 97.5
<i>p</i> -Hydroxybenzaldehyde Syringaldehyde ..	"X222": Benzene 50:50	11.8 1.3	95	4+12	98.0 4.5	2.0 95.5
Syringaldehyde Vanillin	"X222": Ether .. 95:5	7.5 2.5	95	11+50	95.5 5.8	4.5 94.2
Syringaldehyde Vanillin	"X222": <i>iso</i> Propyl ether 95:5 ..	11.3 4.3	95	12+70	97.8 14.0	2.2 86.0
Above separation: best division				10+72	93.8 3.7	6.2 96.3
Maleic acid Fumaric acid	Ethyl ether ..	6.7* 0.67*	99	7+15	†	†

* Collander (1949).

† Precise degree of overlap not determinable (see Fig. 4).

phase and the number of transfers predicted by the curves from the partition coefficients determined by Collander (1949) is shown graphically in Figure 1. Although the precise amount of overlap could not be determined owing to the closeness of their absorption curves, as far as can be judged from the curve the separation was accomplished as predicted.

(b) Estimation of Partition Coefficients from Distribution Results

The usefulness of these curves is not limited to cases where the partition coefficients are accurately known. Approximate values can be used to decide if separation by counter-current distribution is practicable. After a trial and error approach has led to a partial separation, the partition coefficients may be calculated by the formulae (22) or (23) given above, and the number of transfers necessary to effect separation can then be read from the curves.

For many separations discussed in this paper, estimates of the partition coefficients of both solutes involved showed appreciable and as yet unexplained

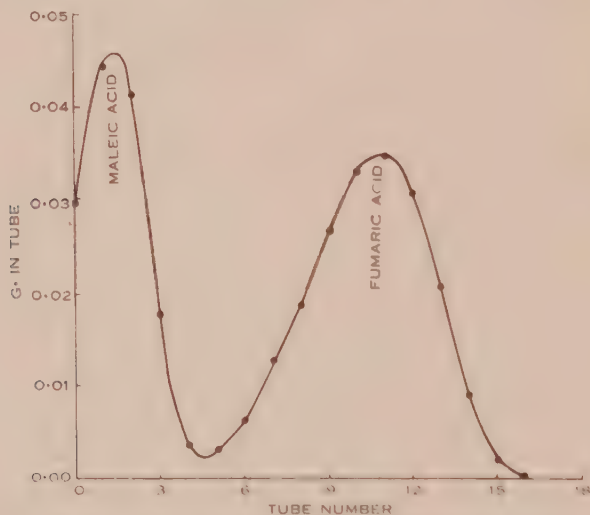


Fig. 4.—Separation of maleic and fumaric acids.
Ether-water, 22 transfers.

departure from measured values. To illustrate the method, the determination for *p*-hydroxybenzaldehyde in the mixture with vanillin is given here. This has been chosen because the observed fractions w_i in each tube give good agreement with theory.

In Table 2, the observed and calculated proportions (w_i and P_i) are set out, together with the values of $\log_e (P_i/w_i)$ required for calculation of an estimate of λ , and hence of the standard error of the estimate of p . The calculations and the estimates obtained are set out at the foot of Table 2. From the value given for the standard error of p' , it may be concluded that the value of p is not likely to differ from p' by more than 0.007.

(c) R_F Values and Partition Coefficients

A valuable extension of this method would be prediction from partition coefficients deduced from observed R_F values. It appeared possible that this could be done by determining the R_F value of a substance with known partition coefficient at the same time as the R_F values of the unknown compounds. That

is the number of transfers necessary for the separation of the substances detected could be predicted from a chromatogram. However, no useful correlation between R_F values and the partition coefficients could be established.

TABLE 2
OBSERVED AND CALCULATED PROPORTIONS FOR A *p*-HYDROXY-
BENZALDEHYDE SEPARATION

Tube No. (<i>r</i>)	$w_r \times 10^3$	P_r (calculated from p') $\times 10^3$	$\text{Log}_e \left(\frac{P_r}{w_r} \right)$
0	10	7	-0.35668
1	46	43	-0.06744
2	118	114	-0.03449
3	189	192	0.01575
4	215	223	0.03654
5	189	193	0.02095
6	126	127	0.00790
7	67	65	-0.03030
8	26	26	0.00000
9	8	8	0.00000
10 } 11 }	6	2	-1.09861
Total	1000	1000	

$$N' = 16, \quad n = 11,$$

$$p' = \frac{\sum r w_r}{N'} = \frac{4.222}{16} = 0.264, \quad \alpha' = \frac{1}{1-p'} = 1.36,$$

$$\sum P_r \log \frac{P_r}{w_r} = 0.00272,$$

$$\frac{1}{\lambda'} = \frac{2 \times 0.00272}{n-1} = 0.000544;$$

$$\lambda' = 1840,$$

$$V(p') = \frac{p'(1-p')}{N\lambda'} = \frac{0.1943}{16 \times 1840} = 6.60 \times 10^{-6},$$

$$\text{Standard error of } p' = 0.00257.$$

This may have been due, at least in part, to the difficulty of working with volatile solvents which necessitated the adoption of a strictly standardized procedure. The time used for conditioning the paper had a pronounced effect; when papers saturated with the aqueous phase were hung on the walls in order to bring the sheet to equilibrium rapidly, erratic R_F values were obtained. The procedure used was chosen so that the conditions could be carefully controlled

and it may be seen from Table 3 that a high degree of reproducibility was attained. Craig and Craig (1950, p. 230) draw attention to lack of correlation between R_F and partition coefficient in many cases and Burma and Banerjee (1950) have produced evidence that the cellulose plays more than an inert role.

TABLE 3
DETERMINATION OF R_F VALUES

Solvent				Vanillin	Syring- aldehyde	<i>p</i> -Hydroxy- benzal- dehyde
		R_F	0.967	0.954	0.428
Benzene	..	Number of determinations	4	4	3
		Standard deviation	0.006	0.007	0.003
		R_F	0.928	0.897	0.648
80% Benzene		Number of determinations	5	5	5
20% " X222 "		Standard deviation	0.006	0.008	0.049
		R_F	0.839	0.797	0.425
50% Benzene		Number of determinations	9	9	9
50% " X222 "		Standard deviation	0.007	0.008	0.048
		R_F	0.737	0.624	0.101
20% Benzene		Number of determinations	7	7	7
80% " X222 "		Standard deviation	0.031	0.052	0.016
		R_F	0.688	0.165	0.022
" X222 "	..	Number of determinations	5	5	5
		Standard deviation	0.025	0.005	0.003

Further development of the theory of partition chromatography may possibly lead to more satisfactory correlation between R_F values and counter-current distribution results.

VI. CONCLUSIONS

The prediction curves developed have led to an elegant and controllable method of separation of the lignin aldehydes which would be applicable on micro- or macro-scale. Although the curves have been tested only with reference to lignin aldehydes and one other pair of substances, their predictions should be generally valid.

VII. ACKNOWLEDGMENT

The authors wish to thank Mr. J. Roberts for the determinations of partition coefficients.

VIII. REFERENCES

- BACHER, J. E. (1951).—Statistical procedures for counter-current distribution and differential spectroscopy. *J. Amer. Chem. Soc.* **73**: 1023.
- BLAND, D. E. (1949).—Separation of vanillin and syringaldehyde by paper partition chromatography. *Nature* **164**: 1093.
- BLAND, D. E., Ho, G., and COHEN, W. E. (1950).—Aromatic aldehydes from the oxidation of some Australian woods and their chromatographic separation. *Aust. J. Sci. Res. A* **3**: 642.
- BURMA, B. P., and BANERJEE, B. (1950).—The role of cellulose in filter paper chromatography. *Science and Culture* **15**: 363.
- COLLANDER, R. (1949).—Die Verteilung organischer Verbindungen zwischen Äther und Wasser. *Acta Chem. Scand.* **3**: 717.
- CRAIG, L. C., and CRAIG, D. (1950).—Extraction and distribution; in "Technique of Organic Chemistry." (Edited by A. Weissberger.) Vol. 3. Chap. IV. (Interscience Publishers Inc.: New York.)
- CREIGHTON, R. H. J., MCCARTHY, J. L., and HIBBERT, H. (1941).—Aromatic aldehydes from plant materials. *J. Amer. Chem. Soc.* **63**: 3049.
- FISHER, R. A., and YATES, F. (1949).—"Statistical Tables for Biological, Agricultural and Medical Research." 3rd Ed. (Oliver and Boyd: London.)
- IDDLES, H. A., LOW, A. W., ROSEN, B. D., and HART, R. T. (1939).—Determination of carbonyl compounds by means of 2,4-dinitrophenylhydrazine. *Industr. Engng. Chem. (Anal. Ed.)* **11**: 102.
- JOHNSON, J. D. A. (1950).—The continuous step-wise extraction of a fixed quantity of a mixture of solutes. *J. Chem. Soc.* **1950**: 1743.
- KENDALL, M. G. (1948).—"The Advanced Theory of Statistics," 4th Ed. Vol. 1. (Griffin: London.)
- NICHOLS, P. (1950).—Useful relations for counter-current distribution computations. *Anal. Chem.* **22**: 915.
- PEARL, I. A., and DICKEY, E. E. (1951).—Separation of syringaldehyde and vanillin by chromatography. *J. Amer. Chem. Soc.* **73**: 863.
- STONE, J. E., and BLUNDELL, M. J. (1951).—A rapid micro method for alkaline nitrobenzene oxidation of lignin and determination of aldehydes. *Anal. Chem.* **23**: 771.

SULPHONAMIDES

III. DISULPHONYL DERIVATIVES OF SOME N-HETEROCYCLIC AMINES

By S. J. ANGYAL,* W. O. MORRIS,* and W. K. WARBURTON*

[Manuscript received October 5, 1951]

Summary

Various disulphonyl derivatives of 2-aminopyridine and 2-aminothiazole are described. It is proved that the two sulphonyl groups are on two different nitrogens; a similar conclusion is reached for dibenzoyl-2-aminopyridine. Several of the disulphonyl compounds exist in two forms which are regarded as geometrical isomers.

I. INTRODUCTION

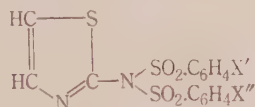
The well-known reaction of 2-aminothiazole with aromatic sulphonyl chlorides in pyridine yields sulphonamides, such as acetylsulphathiazole, but when carried out in other solvents (water, alcohol, acetone) in the presence of inorganic bases, it gives a disulphonyl derivative as the main product of the reaction even when only one equivalent of the sulphonyl chloride is used; with excess sulphonyl halide the yield of the disulphonyl derivative is often nearly quantitative. The patent literature is rich in examples of this reaction (Society of Chemical Industry 1941; Hartmann *et al.* 1945; Chinoïn 1949); and Jensen and Thorsteinsson (1941) obtained a number of such disulphonyl compounds from various sulphonyl chlorides (*p*-acetaminobenzene-, *p*-bromobenzene-, *p*-nitrobenzene-, *p*-toluene-) and several substituted 2-aminothiazoles. Disulphonyl derivatives of 2-aminopyrimidine (Hartmann, von Meyenburg, and Druey 1947), 2-aminoquinazoline (Dewar 1944), and 4-aminoquinoline (Johnson, Woroch, and Buell 1949) are also known. Similar derivatives of 2-aminopyridine have not been described, but it is now shown that with *p*-toluenesulphonyl chloride a *disulphonyl compound* is obtained. This is, however, less stable than the corresponding thiazole derivatives, one of the sulphonyl groups being easily lost on heating with hydroxylic solvents.

An alternative method for the preparation of disulphonyl compounds consists in the reaction of a monosulphonyl derivative with a sulphonyl chloride in the presence of alkali. As one of the sulphonyl groups is readily removed by acid or alkaline hydrolysis, the disulphonyl derivatives can be reconverted into the monosulphonyl compounds. The two types of derivatives are readily separable, as only the latter is soluble in dilute alkali.

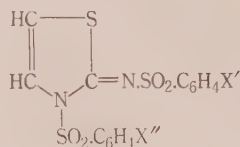
The structure of the monosulphonyl derivatives of 2-aminopyridine, 2-aminothiazole, and 2-aminopyrimidine is established beyond doubt. The sulphonyl group is attached to the amino-nitrogen and not to the ring-atom.

* School of Chemistry, University of Sydney.

The disulphonyl compounds, however, may have either of two isomeric structures, I or II.* The patent literature usually leaves the question of structure undecided but Jensen and Thorsteinsson (1941) favour II because the disulphonyl compounds are unreactive towards methyl iodide.



(I)



(II)

The situation is complicated by the report of Deliwala, Ganapathi, and Shirsat (1943) that bisacetylsulphanilyl-2-aminothiazole was obtained in two forms. The lower melting form (m.p. 128–129 °C.) was converted to the higher one (decomp. about 225 °C.) by prolonged boiling with ethanol. We found that the process can be reversed by dissolving the higher melting form in cold pyridine and precipitating it by the addition of water.† The Indian workers considered the low melting form to have structure I and the higher melting one structure II, but this opinion was not supported by any evidence. Other examples of two forms in a disulphonyl derivative are also known, e.g. bisacetylsulphanilyl-2-amino-4-methylthiazole melts at 150 °C. (Jensen and Thorsteinsson 1941) before and at about 250 °C. (Chinoïn 1949) after boiling in ethanolic solution.

It is now proved, following a suggestion of Brit. Pat. 620,744 (Chinoïn 1949), that the disulphonyl derivatives have structure II. Two different arylsulphonyl groups were introduced into 2-aminothiazole in two different sequences; first acetylsulphathiazole was treated with *p*-toluenesulphonyl chloride, then 2-*p*-toluenesulphonamidothiazole with acetylsulphanilyl chloride. If both groups were on the same nitrogen, i.e. if structure I were correct, the resulting two compounds should have been identical. In fact, however, they were different. On alkaline hydrolysis each disulphonyl compound lost that sulphonyl group which was introduced in the second reaction. This evidence clearly proves structure II. But *each of the two disulphonyl compounds was isolated in two different forms* which were readily interconvertible by the same methods as used with bisacetylsulphanilyl-2-aminothiazole, i.e. by heating in ethanol, and by treatment with cold pyridine.

These two forms can be either dimorphous or isomeric. No difference was noticed in their chemical reactions nor in their ultraviolet absorption spectra. Dimorphism is made unlikely by the observation that the low melting form of

* The referee has suggested that substitution by one sulphonyl group in the 5-position be considered as another possibility, since 2-aminothiazole is readily sulphonated in that position (Hurd and Kharasch 1946). However, disulphonyl derivatives of 5-substituted thiazoles are known (Jensen and Thorsteinsson 1941; Chinoïn 1949). Moreover a sulphonyl group attached to carbon would not be readily hydrolysed by acids and by alkali.

† Jensen and Thorsteinsson (1941) describe a form, m.p. 156 °C., containing 1 mol. of water. This hydrate was not encountered in our work.

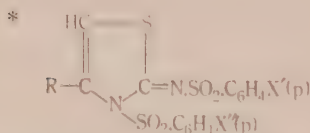
bisacetylsulphanilyl-2-aminothiazole can be recrystallized from ethanol repeatedly and is converted into the higher melting form only by prolonged boiling of the ethanolic solution. (The two mixed disulphonyl compounds change more readily.) Attempts to interchange the two forms by inoculation of their saturated solutions were unsuccessful. We prefer, therefore, to regard the two forms as *geometrical isomers* differing in the steric arrangement around the C—N bond. An undoubted case of such an isomerism, due to the —SO₂—N—C¹—N— grouping, was reported by Angyal and Warburton (1951). It is true that the disulphonyl isomers are readily interconvertible but this may be explained by the fact that the bond order of the bond written as N—C is undoubtedly less than two, because of the strong resonance.

The process of proving structure II was repeated with another pair of sulphonyl groups (*p*-toluene- and *p*-nitrobenzene-) and again two different disulphonyl compounds were obtained.

Table 1 lists the new disulphonyl derivatives prepared either by the direct (method A) or by the stepwise (method B) introduction of sulphonyl groups.

TABLE I.
DISULPHONYL DERIVATIVES.

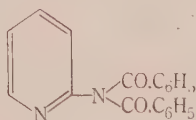
Compound		R	X'	X"	M.P. (°C.)	Found : N (%)	Calc. for : N (%)	Yield (%)		
								Method A di-	mono-	B di-
Disulphonyl derivatives of 2-amino- thiazole*	III	CH ₃	CH ₃	CH ₃	196.5–197 182–187 (decomp.)	6.65 6.55	C ₁₄ H ₁₄ O ₄ S ₂ N ₂ : 6.65	50	31	
	IV	C ₆ H ₅	CH ₃	CH ₃	216	5.9	C ₂₃ H ₁₈ O ₄ S ₂ N ₂ : 5.8	5	4	3
	V	H	CH ₃	NHAc	176–178 149–150	9.35 9.25	C ₁₄ H ₁₁ O ₄ S ₂ N ₂ : 9.3	—	—	80
	VI	H	NHAc	CH ₃	168–169 110–111	9.25 9.35	C ₁₄ H ₁₁ O ₄ S ₂ N ₂ : 9.3	—	—	60
	VII	H	CH ₃	NO ₂	197–198	9.8	C ₁₄ H ₁₃ O ₄ S ₂ N ₂ : 9.55	—	—	70
	VIII	H	NO ₂	CH ₃	167–167.5	9.8	C ₁₄ H ₁₃ O ₄ S ₂ N ₂ : 9.55	—	—	70
Disulphonyl derivative of 2-amino- pyridine	IX	—	CH ₃	CH ₃	176–177 (decomp.) 150–151	7.2 7.1	C ₁₄ H ₁₄ O ₄ S ₂ N ₂ : 6.95	48	30	0



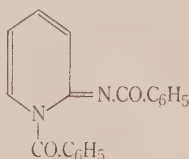
Treatment with pyridine gave a second form of some of the compounds but in those cases where this failed no further attempts were made to prepare isomers. The disubstituted pyridine derivative was also obtained in two forms.

The low reaction rate of 2-amino-4-phenylthiazole is worth commenting upon: by either method the yields were very low even when the time of the reaction was prolonged.

Clarification of the structure of the disulphonyl derivatives suggested an investigation of the diacyl compounds. Dibenzoyl-2-aminopyridine is well known (Tschitschibabin and Bylinkin 1922) and both possible isomeric formulae, X and XI, have been used to describe its structure. In a recent paper, Huntress and Walter (1948) used only X but gave no reason for their choice. It seemed to us that the structure could be established by the same method as that used for the disulphonyl compounds.



(X)



(XI)

However, attempts to benzoylate 2-acetamidopyridine and to acetylate 2-benzamidopyridine were uniformly unsuccessful. It was even found impossible to introduce another benzoyl group into 2-benzamidopyridine by the Schotten-Baumann reaction, which produced the dibenzoyl derivative readily when applied to 2-aminopyridine.

Though the original plan failed, this negative result is regarded as strong evidence in favour of XI. If 2-benzamidopyridine is not attacked, it cannot be an intermediate in the formation of the dibenzoyl compound, and as a simultaneous attack by two molecules of benzoyl chloride is unlikely, it has to be assumed that the ring-nitrogen is benzoylated first. The intermediate thus formed is not very stable and is further benzoylated at a fast rate. Such a sequence of reactions can yield only XI and not X.

II. EXPERIMENTAL

All melting points are corrected.

(a) Preparation of Disulphonyl Compounds, Method A

The appropriate amine (0.1 mol.), the sulphonyl chloride (0.22 mol.), sodium bicarbonate 0.30 mol.), and acetone (100 ml.) were heated under reflux for 1 hour (4 hours for IV). The acetone was removed and water added to precipitate the product. By triturating it in a mortar with 0.5N sodium hydroxide solution (100 ml.) and filtering, it was freed from monosulphonyl compound which was recovered on acidification of the filtrate. The disulphonyl compound was decolorized when necessary with charcoal and recrystallized from ethanol. This gave the high melting form in the yield indicated in Table 1. IX was crystallized from dry benzene; on crystallization from alcohol it changed into 2-*p*-toluenesulphonamidopyridine.

(b) 2-*p*-Toluenesulphonamidopyridine

Obtained as a by-product in the preparation of IX, this compound was crystallized from alcohol, and melted at 215–216 °C.

Found: N, 11.35%.

Calculated for $C_{12}H_{12}O_3SN_2$: N, 11.3%.

(c) *4-Phenyl-2-p-toluenesulphonamidothiazole*

This by-product in the preparation of IV was prepared in 75% yield by the interaction of 2-amino-4-phenylthiazole (2 g.) and *p*-toluenesulphonyl chloride (2 g.) in pyridine (8 ml.) at 25 °C. for 3 days. It crystallized from alcohol in plates containing one mole of solvent which was lost at 100 °C., m.p. 151–151.5 °C. (Loss on drying: 12.2%. Calculated for $C_{16}H_{14}O_2S_2N_2 \cdot C_2H_5OH$: 12.2%.)

Found: N, 8.6%.

Calculated for $C_{16}H_{14}O_2S_2N_2$: 8.5%.

(d) *Preparation of Disulphonyl Compounds, Method B*

To a solution of the appropriate sulphonamidothiazole (0.1 mol.) and sulphonyl chloride (0.11 mol.) in acetone (100 ml.), sodium hydroxide solution (10%; 45 ml.) was added over a period of 10 minutes. After another 10 minutes (1 hour for IV) the solution was diluted with water and the precipitate purified as in method A. In the preparation of IV most of the starting material was recovered.

(e) *Interconversion of Isomers*

Compounds IV, VII, and VIII on solution in cold pyridine (1 g. in 5 ml.) and dilution with water (20 ml.) were unchanged. III, V, VI, and IX under these conditions gave lower melting isomers which were dried at room temperature under reduced pressure. They were all converted to the high melting form on refluxing for an hour in ethanol.

The low melting form of VI, when taken up slowly from 80 °C. in an oil-bath, melted at 167–168 °C., but when placed in the bath at 105 °C. and taken up rapidly, it melted at 110–111 °C., then solidified, and remelted at 160–162 °C. with decomposition. The low melting form of V melted at 149–150 °C. with slight darkening, but melting was not quite complete before resolidification started, and the compound remelted at 176–178 °C. The low melting forms of III and of bisacetylsulphanilyl-2-aminothiazole did not resolidify.

(f) *Absorption Spectra*

Ultraviolet absorption spectra were measured on ethanolic solutions (2 mg. in 100 ml.) with the Beckman spectrophotometer. Compound V had a minimum at 2450 Å (log ϵ 3.90) and maximum at 2770 Å (log ϵ 4.46). VI had minima at 2500 Å (log ϵ 4.19) and 2820 Å (log ϵ 4.01), and maxima at 2380 Å (log ϵ 4.28), 2620 Å (log ϵ 4.25), and 2990 Å (log ϵ 4.14). Bisacetyl-sulphanilyl-2-aminothiazole had a minimum at 2300 Å (log ϵ 3.89) and a maximum at 2670 Å (log ϵ 4.56).

III. ACKNOWLEDGMENTS

The authors wish to thank Monsanto Chemicals (Aust.) Pty. Ltd. for a gift of starting materials. They are also indebted to Mrs. E. Bielski, University of Sydney, for the analyses.

IV. REFERENCES

- ANGYAL, S. J., and WARBURTON, W. K. (1951).—*Aust. J. Sci. Res. A* **4**: 93.
 CHINOIN GYOGYSZER ÉS VEGYÉSZETI TERMÉKEK GYÁRA R. T. (1949).—Brit. Pat. 620,744 (Mar. 30). (*Chem. Abstr.* **43**: 6669 (1949).)
 DELIWALA, C. V., GANAPATHI, K., and SHIRSAT, M. V. (1943).—*Proc. Indian Acad. Sci.* **18A**: 360.
 DEWAR, M. J. S. (1944).—*J. Chem. Soc.* **1944**: 621.
 HARTMANN, M., CERNI, F., DRUEY, J., and VON MEYENBURG, H. (1945).—U.S. Pat. 2,386,852 (Oct. 12). (*Chem. Abstr.* **40**: 5532 (1946).)
 HARTMANN, M., VON MEYENBURG, H., and DRUEY, J. (1947).—U.S. Pat. 2,429,184 (Oct. 14). (*Chem. Abstr.* **42**: 925 (1948).)
 HUNTRESS, E. H., and WALTER, H. C. (1948).—*J. Org. Chem.* **13**: 735.

- HURD, C. D., and KHARASCH, N. (1946).—*J. Amer. Chem. Soc.* **68**: 654.
- JENSEN, K. A., and THORSTEINSSON, Th. (1941).—*Dansk Tidsskr. Farm.* **15**: 41.
- JOHNSON, W. S., WOROCH, E. L., and BUELL, B. G. (1949).—*J. Amer. Chem. Soc.* **71**: 1903.
- SOCIETY OF CHEMICAL INDUSTRY IN BASLE (1941).—Brit. Pat. 533,495 (Feb. 14). (*Chem. Abstr.* **36**: 1050 (1942).)
- TSCHITSCHIBABIN, A. E., and BYLINKIN, J. G. (1922).—*Ber. dtsh. chem. Ges.* **55**: 998.

SULPHONAMIDES

IV. THE REACTION OF N-HETEROCYCLIC AMINES WITH SULPHONYL HALIDES

By S. J. ANGYAL*

[*Manuscript received October 5, 1951*]

Summary

The hypothesis is advanced that the reaction of sulphonyl halides with amino-derivatives of nitrogen-containing cyclic compounds occurs on the ring-nitrogen (which is the more basic), followed either by reaction of the amino-group with another molecule of sulphonyl halide, or by migration of the sulphonyl group from the ring to the amino-nitrogen. Several apparently anomalous cases of this reaction are explained.

I. INTRODUCTION

The discovery of the therapeutic activity of sulphapyridine was followed by a search for more effective analogues, involving mainly the reaction of amino-compounds of the *N*-heterocyclic series with acetylsulphanilyl chloride (for an extensive list see Northey 1948). In most cases the reaction was carried out in the presence of pyridine and yielded the expected acetylsulphanilamido-heterocycle; but sometimes anomalous results were obtained. Thus, in the absence of pyridine, many amines yielded disulphonyl compounds (see Part III of this series, Angyal, Morris, and Warburton 1952). 2-Aminothiazoline, however, gave a disulphonyl derivative in the presence of pyridine (Raiziss and Clemence 1941), and a mono-compound in its absence (Hunter and Kolloff 1943); in the latter derivative the sulphonyl group was shown to have combined not with the amino-group, but with the ring-nitrogen. 2-Amino-oxazoline behaved in the same way (Jensen 1942). 2-Aminobenzimidazole was substituted on one of its ring-nitrogen atoms only (Price and Reitsma 1947). 2-Amino-indole also reacted on the ring-nitrogen (Barber 1945); the sulphanilamido-derivative could only be obtained by a round about way. No explanation has been offered for these anomalous results.

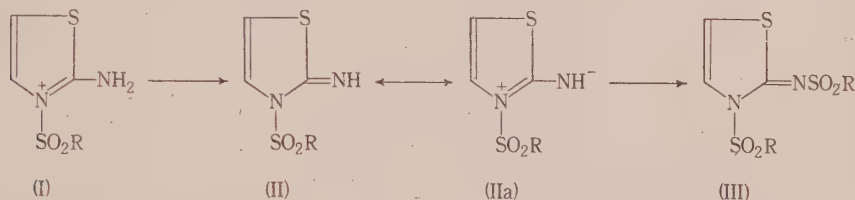
In Part III of this series the theory was advanced that when 2-aminopyridine was benzoylated, the ring nitrogen reacted first. A similar course is now proposed for the reaction with sulphonyl chlorides and it is shown that by this assumption the above-mentioned anomalies can be explained.

II. DISCUSSION

It is well established that when a base of the type of 2- or 4-aminopyridine is converted into its conjugate acid, the proton adds to the ring-nitrogen (Craig and Short 1945; Steck and Ewing 1948). Reaction with alkyl halides also

* School of Chemistry, University of Sydney.

occurs on the ring-nitrogen, again indicating that this is the centre of greater electron availability. As sulphonyl halides presumably react by addition of the lone pair of electrons on a nitrogen atom to the sulphur atom of the sulphonyl group (cf. Shepherd 1947, footnote, p. 276), it is logical to assume that here also the ring-nitrogen would react preferentially. Thus with 2-aminothiazole, the first reaction product would be the ion I which, in the alkaline medium, would lose a proton to yield II. In this compound the imino-nitrogen has a high electron density due to the contribution of the fully aromatic IIa form to its resonance structure and therefore it would react readily with a second molecule of sulphonyl chloride to yield III.



Goldschmidt and Ruttink (1947), studying the reaction of acetylsulphanil chloride with guanidine (which also yields a mono- and a disulphonyl derivative), have proposed a similar course for the reaction; they expressed their belief that the disulphonyl derivative was formed not from the stable acetylsulphaguanidine, but from an unstable isomer, analogous to II.

Unfortunately compounds of the type of II have never been isolated* but several facts support the above given sequence of reactions:

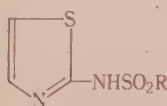
(i) Direct introduction of two sulphonyl groups into a heterocyclic amine often gives a better yield than the reaction of a sulphonamido-heterocycle with a second mole of sulphonyl halide. This indicates that the sulphonamido-compound is not an intermediate in the formation of the disubstituted derivative and therefore the first sulphonyl group did not react with the amino-group. A similar argument is used by Goldschmidt and Ruttink (1947); they point out that acetylsulphaguanidine hardly reacts at all with acetylsulphanil chloride.

(ii) In the absence of pyridine, even if only one mole of a sulphonyl halide is used, considerable amounts of disulphonyl compound are formed. This fact indicates that the second group is introduced at a faster rate into the intermediate than the first into the amine, a behaviour to be expected if II, but not if IV, is the intermediate product.

(iii) In Part III of this series it has been shown that the rate of the reaction of 2-amino-4-phenylthiazole (V) with toluene-*p*-sulphonyl chloride is very low; not only the di-, but the monosubstituted product is slowly formed. Inductive

* A compound described as II ($R = p\text{-AcNH.C}_6\text{H}_4\text{-}$ and with a methyl group in the 4-position) by Bann *et al.* (1944) was later shown by Hoggarth (1947) to be *p*-acetaminophenyl 2-amino-4-methyl-5-thiazyl sulphone.

or resonance effects do not explain this observation; steric hindrance by the phenyl group, however, can account for it provided the rate determining step is a reaction of the ring-nitrogen, and not of the amino-group.



(IV)



(V)



(VI)

The instability of II is due to the attachment of the strongly electron-attracting sulphonyl group to a nitrogen atom which has a low electron density due to the resonance $\text{II} \leftrightarrow \text{IIa}$. Its structure is similar to that of the ion VI, formed from pyridine and sulphonyl halides, which is immediately hydrolysed by water, and which is used as a powerful arylsulphonating agent. The same situation prevails in the disulphonyl derivatives (III); hence the easy hydrolysis, by acids or bases, of the sulphonyl group attached to the ring-nitrogen, whilst the other sulphonyl group is very resistant to these reagents. This also explains a remarkable reaction discovered by the author (Chinoin 1940), namely, that in the presence of pyridine the disulphonyl compounds (III) can transfer one sulphonyl group onto an unsubstituted amine, acting thereby as sulphonating agents. Thus, when equimolecular amounts of 2-aminothiazole and bis-(acetyl-sulphanilyl)-2-aminothiazole (III, $\text{R} = p\text{-AcNH.C}_6\text{H}_4\text{-}$) were heated in pyridine, 2 moles of 2-acetylsulphanilamidothiazole (IV, $\text{R} = p\text{-AcNH.C}_6\text{H}_4\text{-}$) were formed. This reaction is the reason why, in the presence of pyridine, only monosulphonyl compounds are obtained from equimolecular amounts of heterocyclic amines and sulphonyl halides. Any disulphonyl derivative formed in the reaction will lose a sulphonyl group to any unsulphonated amine. When 2 moles of sulphonyl halides are used, however, disulphonyl derivatives can be obtained even in the presence of pyridine (Deliwala, Gunapathi, and Shirsat 1943).

It is assumed that the unstable ring substituted monosulphonyl derivative (II), if not sulphonated further, will rearrange to the stable sulphonamido-compound (IV). This, rather than the direct reaction on the amino-group, would explain the formation of IV in the absence of pyridine. Such rearrangements are known to occur; a well-authenticated case is that of indazole (VII)

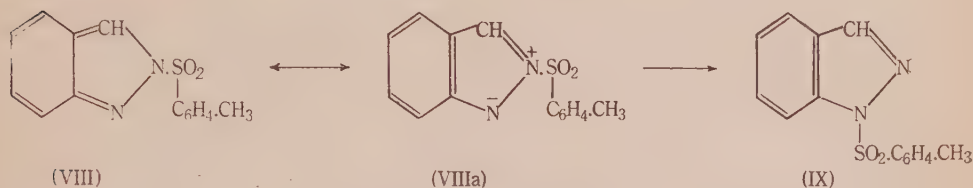


(VII)

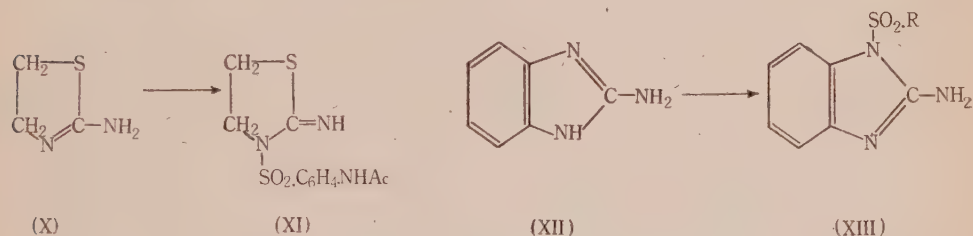
(VIIa)

(Auwers 1925) which reacts with toluene-*p*-sulphonyl chloride (and with other acid halides) in the 2-position where the electron density is higher (due to resonance with VIIa) than on $\text{N}_{(1)}$. The resulting 2-toluene-*p*-sulphonyl-indazole (VIII), because of the electron shift towards the fully aromatic form

VIIIa, has the sulphonyl group attached to a nitrogen of low electron density, and is therefore unstable. On heating, it rearranges to the 1-toluene-*p*-sulphonyl derivative (IX).

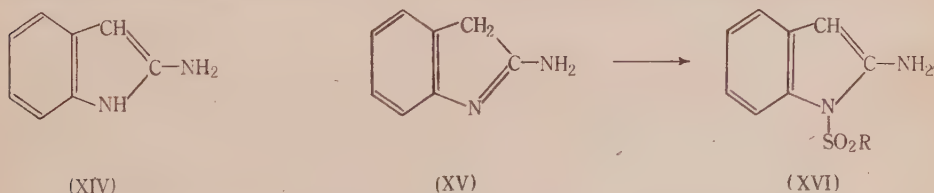


By the above assumption the case of 2-aminothiazoline (X) can be explained as follows : The reaction will again occur on the ring-nitrogen ; loss of a proton will then give 3-acetylsulphanilyl-2-aminothiazoline (XI). Since there is no aromatic resonance in X, no resonance energy has been lost in converting it to XI ; the latter compound is therefore stable. In pyridine, in which the reaction of sulphonyl halides is more vigorous, the imino-group is also sulphonated. The same explanation applies to the case of 2-amino-oxazoline.



Again, 2-aminobenzimidazole (XII) reacts on the most basic N₍₃₎ atom. In this case a proton can be lost from the other ring-nitrogen without loss of aromaticity to give XIII. The ring-substituted derivative is therefore stable. A similar explanation probably applies to the case of 3-aminopyrazole.

The reaction of 2-aminoindole with acetylsulphanilyl chloride is not so readily explained. If aminoindole has structure XIV, reaction on the ring-nitrogen is unexpected because this nitrogen would be very feebly basic. But if aminoindole has the tautomeric structure XV, the case becomes similar to that



of 2-aminobenzimidazole: reaction would occur on the ring-nitrogen and subsequent loss of a proton from the methylene group would yield compound XVI. It could be expected that XV is more stable than XIV because its resonance energy is probably greater. The resonance energy of XIV will not be much larger than that of indole for which the value of 54 kcal./mol. is given

(Pauling 1939, p. 129). On the other hand, XV will have a resonance energy similar to that of *N*-phenylformamidine. This latter value has, apparently, not yet been determined but, since carboxylic acids and carboxamides have similar resonance energies, the values for amidines will probably be of the same order too. A good approximation of the resonance energy of XV may then be the value of 71 kcal. mol. given for formanilide (Wheland 1944). Experiments in progress seem to support structure XV.

III. REFERENCES

- ANGYAL, S. J., MORRIS, W. O., and WARBURTON, W. K. (1952).—*Aust. J. Sci. Res. A* **5** : 368.
AUWERS, K. v. (1925).—*Ber. dtsh. chem. Ges.* **58** : 2081.
BANN, B., KRUG, P., WHEELER, D. E., TAYLOR, W., and GLADDING, G. (1944).—*Brit. Pat.* 559,385 (Feb. 14). (*Brit. Chem. Abstr. B* **2** : 131 (1944).)
BARBER, H. J. (1945).—*J. Amer. Chem. Soc.* **67** : 489.
CHINOIN GYOGYSZER ÉS VEGYÉSZETI TERMÉKEK GYÁRA R. T. (1940).—*Brit. Pat.* 542,160 (Nov. 29). (*Brit. Chem. Abstr. B* **3** : 115 (1942).)
CRAIG, D. P., and SHORT, L. N. (1945).—*J. Chem. Soc.* **1945** : 419.
DELIWALA, C. V., GANAPATHI, K., and SHIRSAT, M. W. (1943).—*Proc. Indian Acad. Sci.* **18A** : 360.
GOLDSCHMIDT, St., and RUTTINK, J. (1947).—*Rec. Trav. Chim. Pays-Bas* **66** : 640.
HOGGARTH, E. (1947).—*J. Chem. Soc.* **1947** : 114.
HUNTER, J. H., and KOLLOFF, H. G. (1943).—*J. Amer. Chem. Soc.* **65** : 156.
JENSEN, K. A. (1942).—*Dansk Tidsskr. Farm.* **16** : 1.
NORTHEY, E. H. (1948).—"The Sulfonamides and Allied Compounds." (Reinhold Publ. Corp. : New York.)
PAULING, L. (1939).—"The Nature of the Chemical Bond." (Cornell Univ. Press.)
PRICE, C. P., and REITSEMA, R. H. (1947).—*J. Org. Chem.* **12** : 269.
RAIZISS, G. W., and CLEMENCE, L. W. (1941).—*J. Amer. Chem. Soc.* **63** : 3124.
SHEPHERD, R. G. (1947).—*J. Org. Chem.* **12** : 275.
STECK, E. A., and EWING, G. W. (1948).—*J. Amer. Chem. Soc.* **70** : 3400.
WHELAND, W. G. (1944).—"Theory of Resonance and its Application to Organic Chemistry." p. 70. (John Wiley and Sons : New York.)

THE CHEMISTRY OF EUCALYPT KINOS

II. AROMADENDRIN, KAEMPFEROL, AND ELLAGIC ACID

W. E. HILLIS*

[Manuscript received January 10, 1952]

Summary

Aromadendrin has been isolated in about 5 per cent. yield from the kino of *Eucalyptus calophylla* and in 1.1–1.3 per cent. yield from the kino of *E. corymbosa*. It has been identified as 3, 4', 5, 7-tetrahydroxyflavanone. Kaempferol was isolated from *E. calophylla* kino and ellagic acid from both kinos.

I. INTRODUCTION

Lauterer (1895) claimed that the characteristic property of the "Turbid" group of kinos (Part I of this series, Hillis 1951) was due to ellagic acid, but later Maiden and Smith (1895) and Smith (1896) showed the turbidity was due to eudesmin and/or aromadendrin and that the latter gave the same colour reactions with nitric acid as did ellagic acid.

The kinos known to contain aromadendrin are from *Eucalyptus hemiphloia* F. Muell. (Maiden and Smith 1895), *E. calophylla* R. Br. (Smith 1896), *E. microcorys* F. Muell., *E. tessellaris* F. Muell., *E. eximia*, Schau., and *E. maculata* Hook., (Smith and Baker 1913). By means of a colour test claimed to be specific for aromadendrin, Ware (1925) concluded that aromadendrin was definitely present in the kinos of *E. calophylla*, *E. coriacea* A. Cunn., *E. gunnii* Hook., *E. eximia*, *E. haemastoma* Sm., *E. leucoxydon* F. Muell., *E. maculata*, *E. punctata* DC., *E. viminalis* Labill., and *Angophora intermedia* DC. Traces were reported to be present in the kinos of some other eucalypt species as well as certain tissues of some *Acacia* species.

Unsuccessful attempts have been made to determine the constitution of aromadendrin (Smith 1896; Nishikawa and Robinson 1922; Phillips 1931).

In the present investigation the kinos of *E. calophylla* and *E. corymbosa* Sm. (syn. *E. gummifera* (Gaertn.) Hochr.) were separated, by means of a liquid-liquid ether extraction, into an ether-soluble red pigment and a brown coloured residue. From the former a light coloured crystalline material was obtained which on crystallization yielded aromadendrin and kaempferol. Ellagic acid separated from the brown coloured residue.

Approximately 5 per cent. of aromadendrin was isolated from a sample of *E. calophylla* kino. Three freshly collected samples of liquid *E. corymbosa* kino yielded 1.3, 1.3, and 1.1 per cent. aromadendrin. The purified material gave an analysis for $C_{15}H_{12}O_6$; four hydroxyl groups were estimated and the

* Division of Forest Products, C.S.I.R.O., Melbourne.

usual colour reactions indicated the structure of a flavanone. Alkali fusion yielded phloroglucinol and alkaline degradation of its dimethyl derivative yielded anisic acid and a material with the same melting point and colour tests as 4-methylphloracetophenone. Aerial oxidation of a solution of aromadendrin in 1N sulphuric acid gave kaempferol.

Aromadendrin was formulated as 3,4',5,7-tetrahydroxyflavanone which structure was confirmed by reducing kaempferol with sodium hydrosulphite at 100 °C. by the method of Pew (1948).

The labile nature of aromadendrin was shown by the formation of 7-methylkaempferol on methylation using dimethyl sulphate and potassium carbonate in acetone. Similar dehydrogenation has been observed with other flavanonols—fustin (Oyamada 1939) and pinobanksin (Linstedt 1950).

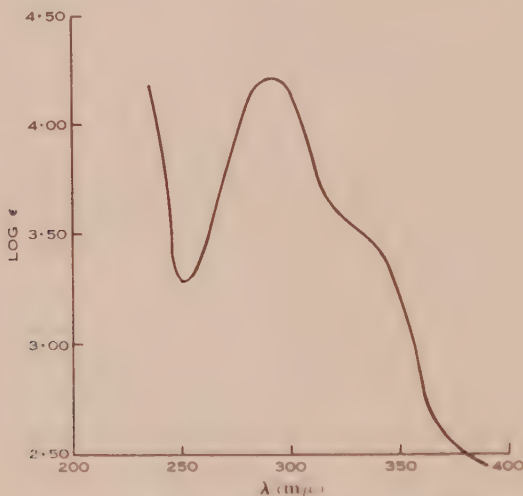


Fig. 1.—Absorption spectrum of aromadendrin.

λ_{max} 292 mμ; $\log \epsilon_{max}$ 4.24.

λ_{min} 250 mμ; $\log \epsilon_{min}$ 3.28.

The ultraviolet absorption spectrum of aromadendrin (Fig. 1) has the same maximum and minimum values as pinobanksin (Linstedt 1950), but the point of inflexion is not so sharp and occurs at a lower $\log \epsilon$ value. An unidentified compound occurred with kaempferol-3-rhamnoside in the powder from a "shake" in the wood of one *Azalia* species (King and Acheson 1950). From the description of its spectrum, melting point, and composition it is probably aromadendrin.

After the completion of this work, a translation of the paper by Ueda, Fukushima, and Kondō (1943) became available. They had isolated from *Cercidiphyllum japonicum* S. et Z., a dihydrokaempferol called katuranin which, when heated, coloured at 200 °C. and decomposed at 224–225 °C. Similar values were obtained with aromadendrin, if the usual soda glass capillary was heated at a very slow rate. If Pyrex capillaries were used, the decomposition point was raised to 237–241 °C., which corresponds to that found by Pew (1948) for

the dihydrokaempferol isolated from *Nothofagus dombeyi* Blume and *Prunus serotina* Ehrh. As it was found that aromadendrin decomposes to kaempferol when kept at its melting point for 1 minute, the melting points were determined at a rapid rate of heating, when values of 247–248.5 °C. were obtained.

The melting point of dimethylaromadendrin (187–187.5 °C.) was similar to that of the dimethylkaturanin (183–184 °C., Uoda, Fukushima, and Kondô, 1943). The author was unable to prepare a crystalline tetra-acetate by their method using sulphuric acid as catalyst or with the aid of perchloric acid or pyridine as catalysts.*

As aromadendrin and katuranin are evidently the same dihydrokaempferol, it is suggested that these trivial names be dropped in favour of dihydrokaempferol.

Chromatographic examination of *E. calophylla* kino revealed the presence of a component with an R_F of 0.67 (Hillis 1951). During the recrystallization of the crude aromadendrin obtained from this kino, a 1 per cent. yield of yellow crystals with the same R_F value was obtained, and this substance was identified as kaempferol. This amount does not represent the total in the kino as chromatographic examination shows that it is also present in the red pigment. The occurrence of flavonol together with its corresponding flavanone has been reported before (Oyamada 1939; Kotaki and Kubota 1940; Linstedt 1950) but this is the first recording of the isolation of kaempferol from wood or a product found in wood.

A fawn compound separated from the brown residue remaining after ether extraction in 0.2 per cent. yield from *E. calophylla* kino and in 0.3 per cent. yield from a *E. corymbosa* kino. It was found also in small amounts in crude aromadendrin and identified as ellagic acid. Chromatographic examination indicates that it is present in all the kinos examined previously (Hillis 1951) and also in the kinos of *E. hemiphloia* F. Muell., *E. macrorrhyncha* F. Muell., and *E. elaeophora* F. Muell. and various tissues of *E. hemiphloia*, *E. macrorrhyncha*, *E. elaeophora*, *E. sideroxylon* A. Cunn., *E. baueriana* Schau. and *E. polyanthemos* Schau. Crystals removed from *E. hemiphloia* heartwood by a micro-manipulator had the same R_F value as ellagic acid. Ware (1925) examined 13 eucalypt kinos by means of colour tests; he devised and detected gallic acid in *E. calophylla* kino but appreciable quantities of ellagitannin and gallotannin in only two other kinos.

II. EXPERIMENTAL

(a) Materials

The sample of *E. calophylla* kino was collected by "bleeding" trees at the Mundaring Weir, Western Australia, and air-drying to about 55% its original weight. When received, the kino was a reddish brown, slightly brittle mass, translucent in 0.5 cm. layers, when it transmitted a brownish ruby colour. The kinos from *E. corymbosa* were collected near Raymond Terrace, N.S.W.

* *Note added in Proof.*—Crystallization was later induced by seeding with crystals of the acetate of dihydrokaempferol, kindly provided by Dr. J. Gripenberg, Finland. After three crystallizations from ether, a colourless substance with melting point 136–136.5 °C. was obtained. Uoda, Fukushima, and Kondô (1943) reported a melting point of 128–129 °C. for katuranin acetate.

(b) *Fractionation of E. calophylla Kino*

Finely powdered air-dry *E. calophylla* kino (1.90 kg., equivalent to 1.69 kg. moisture-free) was added slowly with constant stirring to hot distilled water (2.25 l.). The resulting viscous liquid was extracted in a liquid-liquid extractor with petroleum solvent (b.p. 40–60 °C.) for 36 hours.

The kino syrup was next extracted with ethyl ether at a rate of c. 4 l./hr. for 125 days. The extract was concentrated on a steam-bath until the residue became very viscous, and whilst still hot was dried in a vacuum (70 cm.) oven at 45 °C. for 16 hr., when a porous blood-red mass was obtained—fraction *A*.

After the extraction, the unextractable material was centrifuged, and the precipitate exhaustively extracted with ethanol to yield a fawn crystalline compound—fraction *B*.

The residual part of the kino was dried to a brown solid which was more friable than the original kino.

To fraction *A*, 80% of its weight of absolute ethanol was added, refluxed on a steam-bath for half an hour, then cooled slowly. After 6 weeks the syrup was centrifuged, and the precipitate washed four times with the minimum amount of cold ethanol. The light pink crystals (79.3 g.; 4.7% yield) obtained after drying in a vacuum oven at 45 °C. gave Ware's (1925) colour test for aromadendrin, and is referred to as crude aromadendrin.

Smaller portions of the sample extracted 11 and 16 weeks earlier yielded respectively 5.2 and 5.6% aromadendrin. A similar extraction procedure was followed with the three samples of *E. corymbosa* kino when 1.3, 1.3, and 1.1% yields of aromadendrin were obtained.

(c) *Purification of Aromadendrin*

Crude aromadendrin (obtained from *E. calophylla* kino) was fractionally crystallized from methanol to yield three crystalline substances and a thick black liquor.

Paper partition chromatographic examination with phenol : 2N acetic acid and hydrochloric acid showed that the first crop of crystals (*A 1*) contained mainly a component with R_F 0.41 and an intense mauve fluorescence under ultraviolet light. In these respects it is identical with fraction *B*.

The second crop of crystals (*A 2*) contained mainly a component with R_F 0.84 and an intense greenish fluorescence under ultraviolet light. After a recrystallization from methanol and then 10 crystallizations from absolute ethanol, a product (92% of the crude aromadendrin) was obtained which was found to be homogeneous on chromatographic examination, and was later identified as aromadendrin.

The third crop of crystals (*A 3*) was crystallized eight times from 50% ethanol and yellow needles were obtained in 1% yield of the crude aromadendrin. Chromatographic examination showed that the sample was homogeneous, with an R_F 0.67 and with an intense yellow fluorescence under ultraviolet light.

(d) *Examination of Aromadendrin*

Microelemental analysis by Division of Industrial Chemistry, C.S.I.R.O. All melting points corrected except where otherwise stated.

(i) *Properties*.—The thin colourless needles of aromadendrin were readily soluble in glacial acetic acid, ethyl acetate (3–2% at 8 °C.), and acetone; soluble in boiling water, ethanol (1.7% at 8 °C.), and ethyl ether (1.0% at 5 °C.); slightly soluble in cold water (0.02% at 8 °C.; 0.28% at 25 °C.; 0.66% at 40 °C.), benzene, and chloroform. An aqueous solution gave a yellow precipitate with lead acetate, a transient purple to brown colour with ferric chloride, crimson with concentrated nitric acid, and a yellow colour when warmed with sodium carbonate solution. It reduced hot Fehling's solution and Tollen's reagent; Wilson's boric acid test (1939) was negative. An alcoholic solution treated with hydrochloric acid and zinc or magnesium gave a deep red colour. The material reduced with zinc and acid turned yellow when the solution was neutralized (Ware's (1925) test). When an alcoholic solution was reduced with sodium amalgam, and the supernatant liquor neutralized with hydrochloric acid, a pink colour was obtained.

Aromadendrin could not be sublimed in vacuum. $[\alpha]_D^{25} + 26.5^\circ$ (c. 2% in methanol) $[\alpha]_D^{25} + 51.5^\circ$ (c. 2% in 1:1 acetone and water) (dihydrokaempferol $+45^\circ$, Pew 1948). R_F (phenol:2N acetic acid and hydrochloric acid (1:1)) 0.84 (apple-green fluorescence).

It had m.p. $247-248.5^\circ\text{C}$. in Pyrex capillaries with a rate of heating of $20-240^\circ\text{C}$. in 10 minutes and thereafter at 1°C./min . When the determination was done slowly a m.p. of $237-241^\circ\text{C}$. was obtained. When soda glass capillaries were used, the compound slowly changed colour and shrank. With the same rate of heating, the m.p. was about 15°C . lower than with Pyrex capillaries.

After drying at 78°C . in vacuum for 48 hours, a satisfactory analysis was obtained.

Found: C, 62.5; H, 4.1; OH, 23.8%.

Calculated for $\text{C}_{15}\text{H}_{12}\text{O}_6$: C, 62.5; H, 4.1; OH, 23.6%.

The hydroxyl content was found by determining the residual acetic acid after acetylation with acetic anhydride-pyridine mixture (Hillis 1950).

(ii) *Degradation of Aromadendrin*.—Aromadendrin (0.5 g.) was added to potassium hydroxide (4 g.), sodium hydroxide (4 g.), and water (3 g.) that had been heated to 100°C . in a stainless steel bomb. The mass was well stirred, the bomb closed, and heated at 250°C . for 45 minutes. The brown melt was diluted with water (100 ml.), acidified with 2N sulphuric acid, and extracted with ether (5×30 ml.). The ether extract was shaken with saturated sodium bicarbonate (4×30 ml.) then with water until neutral. The ether was evaporated and the residue crystallized three times from boiling water (once with decolourizing carbon). Yield 0.05 g., m.p. $217-218^\circ\text{C}$., mixed m.p. with phloroglucinol unchanged. No recognizable material could be isolated from the sodium bicarbonate liquors.

(iii) *Oxidation to Kaempferol in Acid Solution*.—Aromadendrin (50 g.) was dissolved in 1N sulphuric acid (800 ml.) and heated under reflux on a steam-bath whilst a slow stream of air was passed over the surface of the solution. The solution became yellow (1.5 hours) and yellow needles precipitated (after 4 hours); after 10 hours the liquid was filtered, the precipitate washed with hot 1N sulphuric acid (3×10 ml.), and the filtrate and washings returned to the flask and reheated as before. The liquor was filtered every 10 hours until no further precipitate appeared. All oxidation products were thoroughly washed with cold water and vacuum-dried at 45°C .

Total yield after 90 hours was 2.6 g. (51% of theoretical), in the first 10 hours 10% (m.p. $270.6-271.6^\circ\text{C}$.) was obtained and successively smaller amounts in each succeeding 10 hours and in each of these cases with a m.p. $262-264^\circ\text{C}$. The yellow crystalline material was crystallized several times from 50% ethanol. A methanol solution gave a green colour with ferric chloride; it did not give a red colour with sodium amalgam, followed by hydrochloric acid, but did so with magnesium and hydrochloric acid, indicating that it was a flavonol. It had m.p. 279°C ., and mixed m.p. with an authentic specimen of kaempferol was unchanged.

Found, after drying at 110°C . in a vacuum: C, 63.0; H, 3.6%.

Calculated for $\text{C}_{15}\text{H}_{10}\text{O}_6$: C, 62.9; H, 3.5%.

The ultraviolet absorption in methanol gave λ_{\max} 2660, 3700 Å; λ_{\min} 2815 Å; $\log \epsilon_{\max}$ 4.38, 4.40, and $\log \epsilon_{\min}$ 3.98 respectively (lit. for kaempferol λ_{\max} 2675; 3700 Å and $\log \epsilon_{\max}$ 4.12, 4.28, Skarzynski 1939). The acetyl derivative melted at $180-181^\circ\text{C}$. (lit. kaempferol tetra-acetate 180°C ., Rao and Seshadri 1946). The kaempferol had an R_F 0.67 (phenol:2N acetic acid and hydrochloric acid) with an intense yellow fluorescence in ultraviolet light.

(iv) *Oxidation to Kaempferol by Fusion*.—Aromadendrin (1.0 g.) was heated in a Pyrex tube at 260°C . for 1 minute. The tube was quickly cooled in cold water and the yellow crystals (0.03 g., m.p. $260-270^\circ\text{C}$.) obtained from the tar had the same R_F value as kaempferol.

(v) *Monomethylation with Dimethyl Sulphate*.—Aromadendrin (1.0 g.), anhydrous acetone (50 ml.), anhydrous potassium carbonate (3 g.), and freshly distilled dimethyl sulphate (0.44 g.) were heated under reflux on a steam-bath for 24 hours. After filtration the potassium carbonate was washed thoroughly with hot acetone and the brown filtrate and washings evaporated to a partly crystalline mass. Yellow crystals (0.28 g.), m.p. $220-221^\circ\text{C}$., were obtained after 6

recrystallizations from 50% ethanol then ethanol. A further 0.08 g. (m.p. 218–219 °C.) was recovered from the potassium carbonate residue after acidification.

The crystals were 7-methylkaempferol (lit. 220–222 °C., Rao and Seshadri 1947) with an acetyl derivative, m.p. 200.5–201 °C. (lit. 200–201 °C., Oesch and Perkin 1914).

(vi) *Methylation by Diazomethane in Ethyl Ether.*—Aromadendrin (1 g.) was dissolved in anhydrous ether (250 ml.) and excess (4 times theoretical) diazomethane in ether was added. After 2 days, the ether was rapidly extracted with saturated aqueous sodium carbonate (2 × 10 ml.), then with water until neutral, and dried with anhydrous sodium sulphate. The ether was distilled and the pale yellow crystalline mass was crystallized four times from 50% ethanol then 75% ethanol to give 0.4 g. of long fine needles with a silky sheen, m.p. 187–187.5 °C. (dimethylkaturanin 183–184 °C., Ueda, Fukushima, and Kondô 1943).

Found: C, 64.5; H, 5.0; CH_3O , 19.6%.

Calculated for $\text{C}_{17}\text{H}_{16}\text{O}_8$: C, 64.5; H, 5.1; $2\text{CH}_3\text{O}$, 19.6%.

The needles of dimethylaromadendrin were slightly soluble in 50% ethanol, ethyl ether, and isopropyl ether, soluble in carbon disulphide and ethyl acetate. When moistened with concentrated nitric acid the dimethyl ether turned a deep blue-green, gave a yellow colour when warmed with aqueous sodium carbonate solution, orange-brown with ferric chloride, and pink with magnesium and hydrochloric acid. Aerial oxidation (q.v.) of a 1N or 5N sulphuric acid solution failed to yield any kaempferol derivatives.

(vii) *Degradation of Dimethylaromadendrin.*—Dimethylaromadendrin (0.5 g.) was added to potassium hydroxide (3 g.) and methanol (30 ml.) contained in a stainless steel bomb, which was then closed and placed in a 100 °C. oven for 30 hours. The reaction mixture was neutralized with 2N sulphuric acid, evaporated, water (50 ml.) added, then extracted with ether (5 × 20 ml.). The ether solution was extracted with saturated sodium bicarbonate (5 × 20 ml.) which was then acidified, extracted with ether, and the extract evaporated to an orange crystalline residue. It was crystallized three times from chloroform: benzene (1:3), giving almost colourless crystals (0.05 g.), m.p. 183.5–184 °C. alone or mixed with an authentic specimen of anisic acid, m.p. 183–184 °C.

The original ether solution was evaporated, the orange residue dissolved in ethanol, and allowed to evaporate slowly when very fine lemon-yellow needles separated. On recrystallization from ethanol: 60–80 °C. petroleum (1:3) a substance (0.006 g.), with m.p. 139–140 °C. was obtained which gave a dark red colour with ferric chloride and deep blue with nitric acid. In these respects it is identical with 4-methylphloracetophenone (Asahina, Shimoda, and Imubase 1927).

(viii) *Acetylation of Aromadendrin.*—Aromadendrin (0.5 g.) was dissolved in pyridine (5 ml.) and acetic anhydride (5 ml.) added. After 24 hours the reaction mixture was worked up in the usual manner to give a colourless gum. No crystalline product could be obtained by this method or when perchloric acid or sulphuric acid was used as catalyst.

(ix) *Racemization of Aromadendrin.*—Aromadendrin (1 g.) was dissolved in 750 ml. of a mixture of equal volumes of concentrated hydrochloric acid and water, refluxed for 5 hours, then diluted to 1500 ml. with water. After settling overnight, the yellow precipitate was filtered and the filtrate extracted with ether (6 × 200 ml.). The ether extract was washed neutral with water (3 × 10 ml.) and the ether distilled to leave a light yellow residue, which was dissolved in boiling water (100 ml.) and allowed to cool overnight. The hexagonal plates (0.5 g.) of a pale tan material were crystallized once from water and decolorizing charcoal, then several times from water: a colourless product was obtained with m.p. 238–240 °C. (Pvrex), no optical activity, and the same R_F value as optically active aromadendrin.

(x) *Synthesis of Aromadendrin.* The reduction method of Pew (1948) using sodium hydro-sulphate was followed and a 33% yield of crude brown product was obtained from kaempferol. This was dissolved in ethanol and the filtered solution allowed to evaporate slowly, when a small amount of colourless plates had formed. These were removed and twice crystallized from 50% ethanol, yield 10%, m.p. and mixed m.p. with racemic aromadendrin 238–240 °C. It had the same R_F value as the natural material.

A further 10% of acemannan was recovered from the mother liquor, by recrystallizing from ethyl acetate-benzene (1:16), isopropyl ether, and water.

Reduction at room temperature by the usual method using excess sodium hydrosulphite (Haas 1936) was unsuccessful.

(iii) *Absorption Spectrum*. A solution (10^{-4} mol. l. purified methanol) was examined in a Beckman Model DU spectrophotometer.

(e) Kaempferol

The yellow crystals (fraction A 3, Section II (c)) were examined and identified as kaempferol as in Section II (d) (iii).

(f) Ellagic Acid

Fraction B was repeatedly crystallized from methanol to give a fawn crystalline compound which on chromatographic examination was shown to be homogeneous, with an R_F of 0.41 (petrol; 2*N* acetic acid; pyridine:acetic acid (1:1)), and an intense mauve fluorescence under ultra-violet light.

No colour was observed on reduction with magnesium and hydrochloric acid, with sodium amalgam or with Wilson's reagents. A green colour was obtained with ferric chloride, a bright red with concentrated nitric acid, and green-yellow with sodium hydroxide. The melting point was higher than 420°C. The substance was recrystallized from pyridine and dried 150°C. in a vacuum.

Found: C, 55.7; H, 2.2%.

Calculated for $C_{14}H_6O_8$: C, 55.6; H, 2.0%.

It was identified as ellagic acid by its tetracarbethoxy-derivative (Feist and Bestehorn 1925; Knebel and Schwan 1942), m.p. 245–247°C. alone or mixed with an authentic specimen of tetracarbethoxyellagic acid.

Found: C, 52.7; H, 3.7%.

Calculated for $C_{24}H_{22}O_{18}$: C, 52.9; H, 3.7%.

The acetyl derivative had m.p. 338–340°C. (corr.) (lit. for tetra-acetyellagic acid 343–346°C., Perkin and Nierenstein 1905).

III. ACKNOWLEDGMENTS

This work was done under the general direction of Dr. W. E. Cohen, Wood Chemistry Section. The samples of *E. corymboza* kino were kindly supplied by Mr. H. W. Morgan and Mr. F. E. Hutchinson of Masonite Corporation (Aust.) Ltd., N.S.W., and the sample of *E. calophyllo* kino by Mr. T. N. Stoate, Conservator of Forests, Western Australia. The author is indebted to Professor F. E. King, University of Nottingham, for the determination of the mixed melting point with an authentic specimen of kaempferol, and to Dr. M. M. Chellaway for the collection of crystals from *E. hemiphloia* heartwood.

IV. REFERENCES

- ARAKIWA, Y., SHINOWA, J., and INUBUSE, M. (1927).—*J. Pharm. Soc. Japan* **550**: 133.
 FEIST, K., and BESTEHORN, H. (1925).—*Arch. Pharm. Berl.* **263**: 16.
 HAAS, E. (1936).—*Biochem. Z.* **285**: 368.
 HILLIS, W. E. (1950).—C.S.I.R.O., Div. For. Prod. Sub-Proc. C6-2, Rep. No. 3.
 HILLIS, W. E. (1951).—*Aust. J. Appl. Sci.* **2**: 385.
 KING, F. E., and ACHESON, R. M. (1950).—*J. Chem. Soc.* **1950**: 168.
 KOTAKI, M., and KUBOTA, T. (1940).—*Liebigs Ann.* **544**: 253.
 LAUTERER, S. (1895).—*Chem. Drugg. Aust.* **10**: 108.

- LINSTEDT, G. (1950).—*Acta Chem. Scand.* **4**: 772.
- MAIDEN, J. H., and SMITH, H. G. (1895).—*J. Roy. Soc. N.S.W.* **29**: 30.
- NISHIKAWA, H., and ROBINSON, R. (1922).—*J. Chem. Soc.* **121**: 839.
- OESCH, J., and PERKIN, A. G. (1914).—*J. Chem. Soc.* **105**: 2352.
- OYAMADA, T. (1939).—*Liebigs Ann.* **538**: 44.
- PERKIN, A. G., and NIERENSTEIN, M. (1905).—*J. Chem. Soc.* **87**: 1412.
- PEW, J. C. (1948).—*J. Amer. Chem. Soc.* **70**: 3031.
- PHILLIPS, L. W. (1931).—*J. Roy. Soc. W. Aust.* **18**: vii.
- RAO, K. V., and SESHADRI, T. R. (1947).—*J. Chem. Soc.* **1947**: 122.
- RAO, P. R., and SESHADRI, T. R. (1946).—*Proc. Ind. Acad. Sci.* **24A**: 456.
- REICHEL, L., and SCHWAB, A. (1942).—*Liebigs Ann.* **550**: 152.
- SKARZYNSKI, B. (1939).—*Biochem. Z.* **301**: 150.
- SMITH, H. G. (1896).—*J. Roy. Soc. N.S.W.* **30**: 135.
- SMITH, H. G., and BAKER, R. T. (1913).—*Aust. Ass. Advanc. Sci.* **14**: 83.
- UODA, H., FUKUSHIMA, B., and KONDŌ, T. (1943).—*J. Agric. Soc. Japan* **19**: 467.
- WARE, A. H. (1925).—*Pharm. J.* 4S, **61**: 131.
- WILSON, C. W. (1939).—*J. Amer. Chem. Soc.* **61**: 2303.

ALKALOIDS OF THE AUSTRALIAN RUTACEAE: *PENTACERAS AUSTRALIS* HOOK. F.

I. ISOLATION OF THE ALKALOIDS AND IDENTIFICATION OF CANTHIN-6-ONE

By H. F. HAYNES,* EVA R. NELSON,† and J. R. PRICE*

[Manuscript received December 5, 1951]

Summary

Three alkaloids, with molecular formulae $C_{14}H_8ON_2$, $C_{15}H_{10}O_2N_2$, and $C_{15}H_{10}OSN_2$, have been isolated from the rain-forest tree *Pentaceras australis* Hook.f. All three are present in the bark of mature trees, the total yield of alkaloids varying from c. 0.2 per cent. (branch bark) to c. 1 per cent. (root bark). The wood contains 0.2 per cent. of a mixture of two of the alkaloids, $C_{14}H_8ON_2$ and $C_{15}H_{10}OSN_2$, while the leaves contain less than 0.01 per cent. of a mixture in which the alkaloids $C_{14}H_8ON_2$ and $C_{15}H_{10}O_2N_2$ are present.

The alkaloid $C_{14}H_8ON_2$ is shown to be representative of a ring system not previously found in the plant kingdom. It is canthin-6-one (II). The principal reactions on which this structure is based are (a) oxidation of the alkaloid to β -carboline-1-carboxylic acid and (b) the reversible opening of a lactam ring with alkali to give *cis*-2-(1'- β -carboly) acrylic acid (III) which may be transformed to the *trans*-isomer from which the alkaloid is not regenerated. Other reactions of the alkaloid support the canthinone structure. Among the compounds described is a dihydro-derivative (XI) of the parent base canthine.

I. INTRODUCTION

Pentaceras australis Hook.f., a rain-forest tree belonging to the tribe Zanthoxyloae of the family Rutaceae (Engler and Prantl 1931), occurs in eastern New South Wales and Queensland and is the only member of the genus. The bark has been found to contain three alkaloids possessing the molecular formulae $C_{14}H_8ON_2$, $C_{15}H_{10}O_2N_2$, and $C_{15}H_{10}OSN_2$. The wood contains two of the alkaloids, $C_{14}H_8ON_2$ and $C_{15}H_{10}OSN_2$, while the leaves contain less than 0.01 per cent. of a mixture consisting of the alkaloids $C_{14}H_8ON_2$, $C_{15}H_{10}O_2N_2$, and a third base which gives a red picrate, but, being present in very small amount, was not further examined. The yield of alkaloids in the bark varies with the age of the tree and the part of the tree from which the bark is obtained (see Table 1). The reactions of the alkaloid $C_{14}H_8ON_2$ have been studied and its structure established; the structures of the alkaloids $C_{15}H_{10}O_2N_2$ and $C_{15}H_{10}OSN_2$ will be dealt with in subsequent papers.

II. STRUCTURE OF THE ALKALOID $C_{14}H_8ON_2$

The alkaloid is optically inactive, contains neither methoxyl nor methylimino-group, cannot be acetylated, and gives no reaction for a carbonyl

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

† Chemistry Department, University of Melbourne; present address: Nicholas Pty. Ltd., Melbourne.

group. It is a monoacid base, forming a *hydrochloride* $C_{14}H_8ON_2 \cdot HCl$, a *picrate* $C_{14}H_8ON_2 \cdot C_6H_3O_7N_3$, and a *methiodide* $C_{14}H_8ON_2 \cdot CH_3I$. The alkaloid is rapidly

TABLE I
YIELD OF ALKALOIDS

Source	$C_{14}H_8ON_2$ (%)	$C_{15}H_{10}O_2N_2$ (%)	$C_{15}H_{10}OSN_2$ (%)
Stem bark (1)* ..	0.14	0.12	0.46
Stem bark (2) ..	0.25	0.12	0.29
Root bark ..	0.15	0.21	0.63
Branch bark ..	0.05	0.04	0.09
Sapling bark ..	—	<0.01	0.01
Wood	0.17	—	0.05
Leaves	<0.01	<0.01	—

* Samples (1) and (2) were from different trees.

oxidized by potassium permanganate in acetone solution to a pale yellow acid, $C_{12}H_8O_2N_2$, which can be decarboxylated to a colourless base $C_{11}H_8N_2$. Comparison of the base and its picrate with authentic specimens establishes the

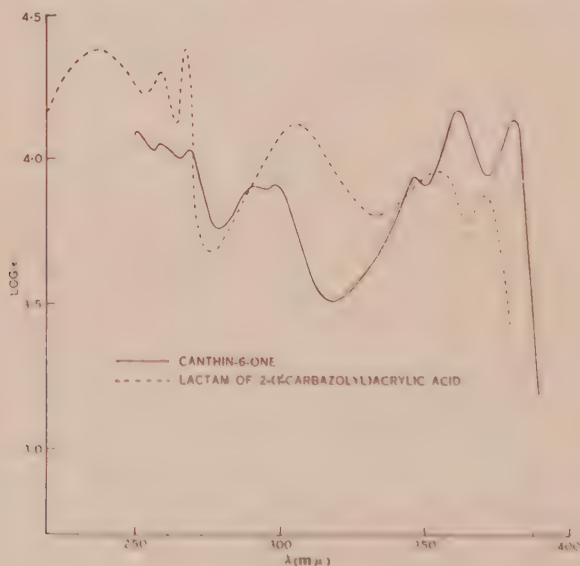
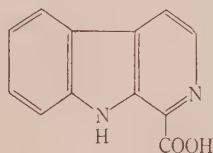


Fig. 1

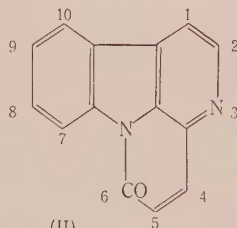
identity of the decarboxylation product as β -carboline. Comparison of its methyl ester with a synthetic specimen shows the acid to be β -carboline-1-carboxylic acid* (I).

* The system of numbering is that recommended in "The Ring Index" (Patterson and Capell 1940).

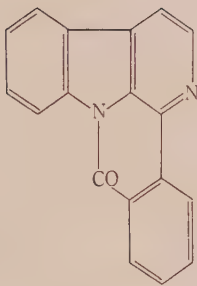
The alkaloid dissolves in boiling aqueous-ethanolic alkali and acidification of the solution gives the hydrate of a yellow acid $C_{14}H_{10}O_2N_2 \cdot H_2O$. This acid is reconverted to the alkaloid under mild alkaline conditions; reversion also takes place during attempted esterification and, slowly, in hot acetic acid although the acid may be recrystallized from this solvent. The behaviour with alkali implies the presence in the molecule of a lactam ring, and, together with the formation of β -carboline-1-carboxylic acid, leads to the conclusion that the alkaloid possesses structure II, alkaline hydrolysis of the lactam giving 2-(1'- β -carboly)acrylic acid (III). The ring system represented by II has not previously been encountered among plant alkaloids, but Manske and Marion (1938) by heating calycanthine with phthalic anhydride obtained a substance to which



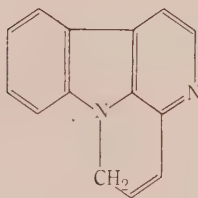
(I)



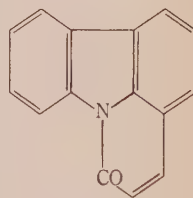
(II)



(V)



(VI)



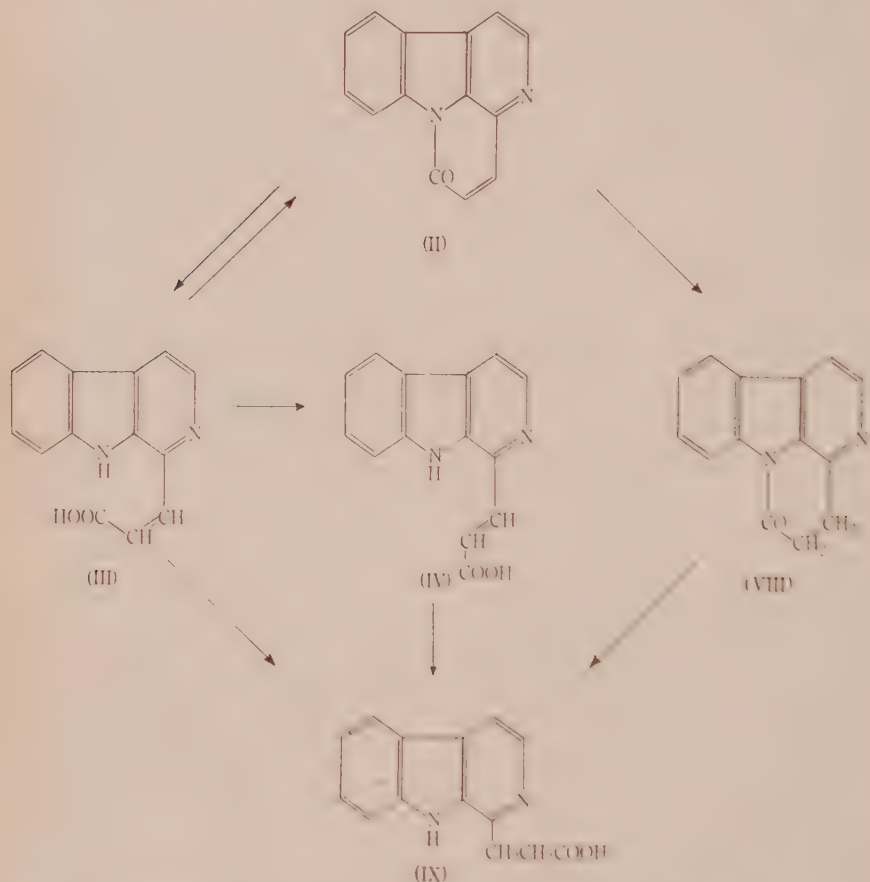
(VII)

they ascribed structure V and for which they proposed the name 12,13-benz-canthin-11-one, regarding it as a derivative of the hypothetical base canthine (VI). Using the system of numbering recommended in "The Ring Index" in preference to that of Manske and Marion (1938), the alkaloid $C_{14}H_8ON_2$ is *canthin-6-one*. The ultraviolet absorption spectrum of canthinone (Fig. 1 and Table 2) closely resembles that of the carbazole analogue (VII) synthesized by Prelog, Szpilfogel, and Battegay (1947), from whose paper the absorption curve in Figure 1 is reproduced.

Prolonged heating of the β -carbolyacrylic acid with alkali transforms it to an isomeric higher melting yellow acid which was isolated as the hygroscopic hemihydrate $C_{14}H_{10}O_2N_2 \cdot \frac{1}{2}H_2O$. Both acids give the same dihydro-acid (IX)

on reduction and the second is evidently the *trans*-isomer.* In agreement with this view the *trans*- β -carbolylarylic acid does not cyclize to canthinone and is easily esterified. The methyl ester, $C_{15}H_{12}O_2N_2$, has been prepared and its ultraviolet absorption spectrum is recorded in Figure 2.

The 4,5-double bond in canthinone is readily hydrogenated in the presence of Raney nickel, giving 4,5-dihydrocanthin-6-one (VIII) which may also be



prepared by treatment of the alkaloid with zinc and acetic acid. This dihydrocanthinone is hydrolysed by alkali to the same 2-(4,5-dihydro-2H-pyrido[1,2-b:3',2'-d]pyridin-6-yl)propionic acid (IX) obtained from both *cis*- and *trans*- β -carbolylarylic acid by means of Raney alloy and sodium hydroxide. The methyl ester of IX may be prepared by direct esterification of the acid or from canthinone by boiling with excess Raney nickel in methanol, hydrogenation and methanolysis both occurring. Likewise the ethyl ester of IX is obtained from canthinone using Adams platinum catalyst

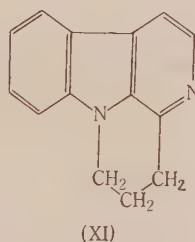
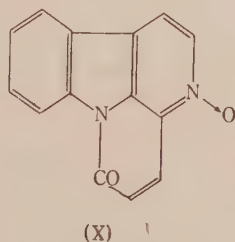
* Cannon, Hughes, and Ritchie (personal communication), in the course of an examination of the alkaloids of *Zanthoxylum suberosum* C. T. White, also observed the formation of this second acid and suggested that the two acids were *cis-trans*-isomers.

in ethanol. Dehydrogenation of β -carbolypropionic acid with selenium dioxide gives a mixture of *trans*- β -carbolyacrylic acid (IV) and canthinone (II), the latter clearly resulting from cyclization of the *cis*-acid. Attempts to recycle β -carbolypropionic acid to dihydrocanthinone were unsuccessful. The ultraviolet absorption spectrum of dihydrocanthinone is shown in Figure 2.

TABLE 2
ULTRAVIOLET ABSORPTION DATA

Substance	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$-\log \epsilon_{max}$
Canthin-6-one	381	4.14	299	3.91	269	4.03
	362	4.17	~ 293	3.90	259	4.05
	347	3.94			251	4.09
4,5 - Dihydrocanthin - 6-one	328	4.00	283	4.09	273	4.05
	315	3.88			261	4.17
Methyl ester of <i>trans</i> - β -carbolyacrylic acid	382	3.94	297	4.30	~ 240	4.33

Canthinone takes up one atom of oxygen when treated with hydrogen peroxide in acetic acid. The product, $C_{14}H_8O_2N_2$, is presumably the N-oxide (X). It is unaffected by sulphur dioxide but this appears to be characteristic of pyridine-N-oxides (see, for example, Ochiai, Ishikawa, and Katada 1943). Zinc and acetic acid remove the added oxygen atom but also reduce the 4,5-double bond giving dihydrocanthinone. Dihydrocanthine (XI) is produced by reducing



canthinone with zinc and hydrochloric acid, a facile reduction of $>CO$ to $>CH_2$ taking place together with the anticipated reduction of the double bond. The hygroscopic liquid base has not been analysed, but it forms a crystalline *hydrate*, a *picrate*, and *picrolonate*, the analyses of which establish the formula $C_{14}H_{12}N_2$. The base is stable towards boiling ethanolic alkali and, as would be expected, reacts with benzaldehyde forming a benzylidene derivative $C_{21}H_{16}N_2$.

The hexahydrocanthinone (XIII) has been synthesized by Hahn and Hansel (1938) by condensing tryptamine hydrochloride with α -ketoglutaric acid and effecting ring closure and decarboxylation of the resulting dicarboxylic

acid (XII) with methanolic hydrogen chloride. Hydrogenation experiments with canthinone, carried out with the object of relating it to Hahn and Hansel's base, led to the uptake of only one mole of hydrogen. Consequently attempts were made to dehydrogenate Hahn and Hansel's base. However, it was

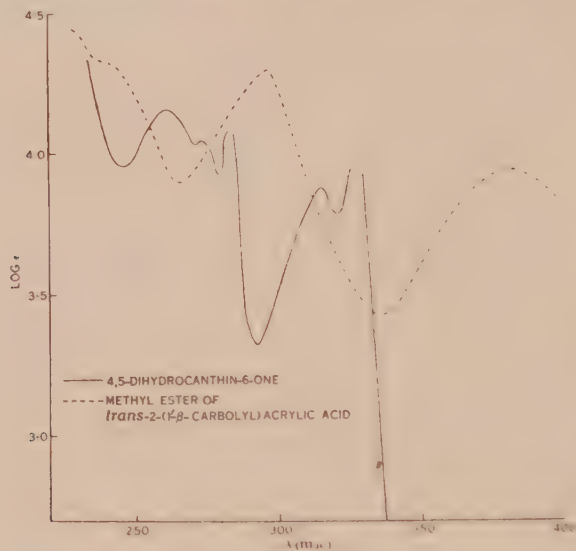
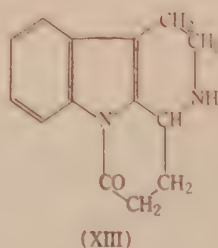
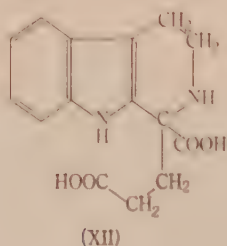


Fig. 2

recovered unchanged after treatment with chloranil, palladized charcoal in the presence of maleic acid, and palladium black in the presence of either maleic or cinnamic acid, procedures which give satisfactory results in the dehydrogenation of either tetrahydrocarbazole or tetrahydroharman. Dehydrogenation of the corresponding carbazole derivative has been achieved by Prelog, Szpilloegel, and Battegay (1947) by means of chloranil, the product being the base VII. Dichromate oxidation either of the acid XII or of the base XIII was similarly unsuccessful.



Tests carried out by Associate Professor F. H. Shaw, Department of Physiology, University of Melbourne, show that canthin-6-one and dihydrocanthine have no noteworthy physiological effects on mice or toads, but that canthinone methiodide exhibits weak curariform activity.

III. EXPERIMENTAL

All melting points are corrected except where otherwise stated. Microanalyses were carried out by Mr. H. L. Oates.

(a) *Extraction of the Bark*

Extraction of the milled bark* was carried out either directly with hot methanol (soxhlet) or, after liming, by percolation with trichloroethylene at room temperature. The latter was much more satisfactory since methanol extracted a great quantity of gummy material which hampered the separation of the alkaloids. In some cases where methanolic extraction was employed, the residue after removal of the methanol was triturated with a thin paste of sodium bicarbonate and water, and extracted with chloroform. After evaporation of the chloroform the residue was treated as in the case of a lime/trichloroethylene extract. Other methanolic extracts, after removal of the solvent, were boiled repeatedly with 2% and finally with 15% aqueous hydrochloric acid. After recovery of the crude alkaloids from the two series of acid extracts they were worked up separately, but along the same lines as the trichloroethylene extracts. The separation of the alkaloids was tedious but straightforward and, though no procedure was strictly adhered to, one example should suffice to illustrate the methods employed. For simplicity, the three alkaloids will be referred to in the description of the extraction as *A* ($C_{14}H_8ON_2$), *B* ($C_{15}H_{10}O_2N_2$), and *C* ($C_{15}H_{10}OSN_2$), respectively.

One kilogram of milled bark (stem) was limed (300 g. lime) and percolated with trichloroethylene until only a negligible amount of material was being removed. The combined percolates were distilled, leaving a brown semicrystalline residue which was heated with a small volume of methanol. After cooling, the almost colourless solid was filtered from the methanol solution and the two fractions worked up separately.

(i) The solid, insoluble in cold methanol, was extracted three times with boiling 2% hydrochloric acid and the acid extracts combined with the methanol-soluble fraction (ii). The acid-insoluble material was washed with water and crystallized from chloroform-ethanol giving 2.66 g. alkaloid *C*. A further 1.15 g. was obtained by concentrating the mother liquors. The final filtrate was evaporated to dryness, boiled with 10% hydrochloric acid, and some acid-insoluble material discarded. The acid solution was basified with ammonia, extracted with chloroform, and the extract set aside for further treatment.

(ii) The methanol solution was evaporated leaving a dark gum which was boiled three times with 2% and finally with 10% hydrochloric acid. These acid solutions were combined with the 2% hydrochloric acid extract from (i) above, and shaken with chloroform to remove a "weakly basic" fraction. They were then basified with ammonia and again shaken with chloroform to remove a "strongly basic" fraction. Evaporation of the solvent from the weakly basic fraction and treatment of the gummy residue with methanol gave a further 0.54 g. alkaloid *C* undissolved by cold methanol. The methanol filtrate was set aside for further treatment. The strongly basic alkaloid fraction was left as a brown crystalline residue after evaporation of the chloroform. Crystallization from chloroform-ethanol gave 0.95 g. alkaloid *B*.

(iii) The mother liquors from alkaloid *B* were combined with the methanol-soluble portion of the weakly basic fraction and with the material extracted from the mother liquors of (i) above. The solvents were evaporated, the residue dissolved in benzene, and the solution extracted successively with 0.25, 0.5, 2, and 5% hydrochloric acid. The benzene solution, after washing with water, was evaporated to dryness and the gummy residue treated with methanol giving 0.20 g. alkaloid *C*. The base recovered from the 0.25% acid extracts was crystallized from chloroform-ethanol giving 0.26 g. alkaloid *B* and 0.30 g. of the more soluble alkaloid *A*. From the 0.5% acid extracts the recovered base gave 0.26 g. alkaloid *A* crystallized from methanol, while from the 2% acid extracts were obtained after crystallization, 0.38 g. alkaloid *A*. The

* All bark, wood, and leaves examined had previously been dried in a current of air at a temperature not exceeding 60°C. and then allowed to equilibrate in air. No allowance has been made for the moisture content of this equilibrated material in the figures quoted for alkaloid content.

mother liquors from the crystallization of the alkaloids from the 0.25, 0.5, and 2% acid extracts were combined with the small amount of alkaloid recovered from the 5% acid extracts and treated with alcoholic picric acid. The precipitated picrate was filtered, washed, and the base regenerated, giving after crystallization from methanol 0.44 g. alkaloid A. The total yield was as follows:

Alkaloid A ($C_{14}H_8ON_2$): 1.38 g., 0.14%

Alkaloid B ($C_{15}H_{10}O_2N_2$): 1.21 g., 0.12%

Alkaloid C ($C_{15}H_{10}OSN_2$): 4.55 g., 0.46%.

Alkaloid A crystallized from methanol in colourless needles, m.p. 162.5–163.5°C.

Found: C, 76.6; H, 3.8; N, 12.6%; mol. wt. (Rast) 234.

Calculated for $C_{14}H_8ON_2$: C, 76.4; H, 3.6; N, 12.7%; mol. wt. 220.

Alkaloid B crystallized from chloroform-ethanol as long thin colourless needles, m.p. 241.5–242°C.

Found: C, 72.1; H, 4.1; N, 11.0%; mol. wt. (Signer) 254.

Calculated for $C_{15}H_{10}O_2N_2$: C, 72.0; H, 4.0; N, 11.2%; mol. wt. 250.

Alkaloid C crystallized from chloroform-ethanol as long colourless needles, m.p. 252.5–253.5°C.

Found: C, 67.5; H, 3.8; S, 11.8; N, 10.5%; mol. wt. (Signer) 284.

Calculated for $C_{15}H_{10}OSN_2$: C, 67.7; H, 3.8; S, 12.0; N, 10.5%; mol. wt. 266.

(b) Extraction of the Wood

Sawdust (900 g.) was extracted with methanol (soxhlet) and the methanol distilled off leaving a dark brown gummy residue which was triturated with a thin paste of sodium bicarbonate and water, and extracted with chloroform. The chloroform solution was evaporated to dryness and the residue extracted three times with 2 l. of boiling 10% hydrochloric acid. The combined acid extracts were diluted with an equal volume of water, shaken with chloroform giving a weakly basic fraction, then basified with ammonia and again shaken with chloroform to give a strongly basic fraction. Evaporation of the chloroform from the strongly basic fraction and crystallization of the residue from methanol gave 1.27 g. of the alkaloid $C_{14}H_8ON_2$, identified by its m.p. and mixed m.p. The residue after evaporation of the solvent from the weakly basic fraction was treated with methanol leaving 0.43 g. of the sparingly soluble alkaloid $C_{15}H_{10}OSN_2$ (m.p. and mixed m.p.). The methanolic filtrate and the mother liquors from the crystallization of the first alkaloid were combined and treated with ethanolic picric acid. Regeneration of the base from the precipitated picrate gave, after crystallization from methanol, a further 0.25 g. of the alkaloid $C_{14}H_8ON_2$, making a total yield of this alkaloid of 1.52 g., that is 0.17%, while that of the alkaloid $C_{15}H_{10}OSN_2$ was 0.43 g. or 0.05%.

(c) Extraction of the Leaves

Milled leaves (1375 g.) were extracted with methanol (soxhlet), the solvent distilled off, and the residue boiled twice with 2 l. of 2% hydrochloric acid. The acid extracts were shaken with chloroform, removing a weakly basic fraction, then basified with ammonia, and again shaken with chloroform to remove a strongly basic fraction. The latter after removal of the solvent was again extracted with hot 2% hydrochloric acid and the acid solution basified and extracted with chloroform. Evaporation of the chloroform solution left a dark brown semi-crystalline residue which was converted to its picrate. Regeneration of the base gave 45 mg. which was dissolved in benzene and chromatographed on alumina. The main fraction (25 mg.) was crystallized from chloroform-ethanol and identified as the alkaloid $C_{14}H_8ON_2$, m.p. and mixed m.p., yield c. 0.002%. The weakly basic fraction after evaporation of the solvent was extracted with hot 5% hydrochloric acid and the recovered base converted to its picrate. Regeneration of the base from the picrate gave 80 mg. which was chromatographed in benzene on alumina. The main fraction (65 mg.) was still not homogeneous; it was again converted to its picrate from which the base was regenerated. This base, after crystallization from light petroleum, was identified as the alkaloid $C_{14}H_8ON_2$ (m.p. and mixed m.p.), yield c. 0.004%. Elution of the

alumina column with chloroform gave c. 10 mg. of an orange crystalline solid giving a dark red picrate melting with decomposition at 270–272 °C. (uncorr.). Owing to the very small yield this base was not further investigated.

(d) *Properties of the Alkaloid C₁₄H₈ON₂ (II)*

The alkaloid C₁₄H₈ON₂ is very soluble in chloroform, may be crystallized conveniently from benzene or methanol, but is almost insoluble in light petroleum and in water. Analysis shows the absence of methoxyl or methylimino-groups, and a solution in chloroform (c, 1.55) was optically inactive. Though only sparingly soluble in 0.01N HCl, the alkaloid dissolves in 0.1N HCl giving a pale yellow solution. The *hydrochloride* crystallized from 1N HCl as pale yellow plates or needles. A specimen dried in a vacuum over P₂O₅ at room temperature shrank and darkened from 220 °C. and melted with decomposition at 244–246 °C. (uncorr.).

Found: Cl, 13.4%.

Calculated for C₁₄H₈ON₂·HCl: Cl, 13.8%.

The hydrochloride dissolves in water but on standing the yellow solution soon deposits colourless crystals of the free base. The *picrate* separates from methanol, in which it is only sparingly soluble, in flat yellow needles, m.p. 262–264 °C. (uncorr.) after softening from c. 240 °C. upwards.

Found: C, 53.9; H, 2.5; N, 15.7%.

Calculated for C₁₄H₈ON₂·C₆H₃O₇·N₃: C, 53.5; H, 2.4; N, 15.6%.

The *methiodide* was prepared in 90% yield by refluxing a solution of the alkaloid in chloroform with excess methyl iodide for 15 hours. The solid which separated crystallized from water as orange-red plates, m.p. 271–273 °C. (uncorr.).

Found: C, 50.0; H, 3.0%.

Calculated for C₁₅H₁₁ON₂I: C, 49.7; H, 3.0%.

The alkaloid gave no colour with ferric chloride in ethanolic solution. It was recovered unchanged (m.p. and mixed m.p.) after (i) refluxing for 4 hours with a mixture of 46% aqueous hydrobromic acid (1 vol.) and acetic acid (2 vols.), and (ii) attempted acetylation with boiling acetic anhydride and pyridine. After refluxing with alcoholic 2,4-dinitrophenylhydrazine hydrochloride and cooling, dull yellow crystals separated. These were evidently an addition compound; the alkaloid was recovered and identified by m.p. and mixed m.p. after decomposition of the 2,4-dinitrophenylhydrazine with cold alkali.

(e) *Oxidation with Permanganate*

Oxidation of the alkaloid by permanganate in acetone solution took place rapidly at room temperature. When oxidation was complete, the precipitated manganese dioxide, suspended in water, was treated with sulphur dioxide giving in good yield an acid which crystallized from acetic acid in yellow needles, m.p. 239.5 °C. (decomp.).

Found: C, 67.7; H, 3.8; N, 12.9%.

Calculated for C₁₂H₈O₂N₂: C, 67.9; H, 3.8; N, 13.2%.

A portion of this acid was esterified with methanolic hydrogen chloride and the resulting ester crystallized first from benzene and then from aqueous methanol. It was obtained as colourless needles, m.p. 168–168.5 °C. not depressed by admixture with a synthetic specimen (see below) of the methyl ester of β-carboline-1-carboxylic acid. Another portion of the acid was decarboxylated by heating in dibutyl phthalate at 240–250 °C. for 1 minute. The solution was cooled, diluted with chloroform, filtered, and extracted three times with 1% hydrochloric acid. The combined acid extracts were basified with ammonia and the precipitated base shaken out with chloroform. Evaporation of the chloroform and crystallization of the residue from benzene gave colourless needles, m.p. 202–203 °C. not depressed by admixture with authentic β-carboline.

Found: C, 78.8; H, 4.8; N, 16.6%.

Calculated for C₁₁H₈N₂: C, 78.6; H, 4.8; N, 16.7%.

The base formed a picrate, yellow needles from methanol, which on standing in contact with the solution slowly changed to orange prisms. Recrystallization again gave yellow needles and both forms melted with decomposition at 264–265 °C. (uncorr.); Kernack, Perkin, and Robinson (1921) record 260 °C. with decomposition for the m.p. of β -carboline picrate.

Specimens of β -carboline-1-carboxylic acid, of its methyl ester and of β -carboline, required for comparison, were prepared from benzalharman which was oxidized by permanganate in acetone solution. The resulting β -carboline-1-carboxylic acid crystallized from acetic acid as yellow needles, m.p. 237.5–238 °C. (decomp.); cf. Snyder, Walker, and Werber (1949) who state that the insolubility of β -carboline-1-carboxylic acid in all solvents tried made crystallization impossible. The synthetic methyl ester melted at 166–166.5 °C.

(f) *Action of Alkali on the Alkaloid $C_{14}H_8ON_2$*

(i) *cis- β -Carbolylacrylic Acid (III)*. The alkaloid is not soluble in cold aqueous alkali but dissolves slowly on heating to give a yellow solution. Solution may be effected more rapidly by adding a concentrated solution of the alkaloid in methanol to hot alkali and boiling off the methanol. Using 10% sodium hydroxide, there separated almost at once from the hot solution a mass of colourless plates consisting of the sodium salt of an acid which was isolated by dissolving the salt in cold water, filtering quickly, and acidifying immediately with acetic acid. The crystalline precipitate of *cis*-2-(1'- β -carbolyl)acrylic acid (III) was recrystallized from acetic acid from which it separated as yellow needles, m.p. 212–214 °C. (decomp.).

Found (on an air-dried specimen): C, 65.6; H, 4.5%.

Calculated for $C_{14}H_{10}O_2N_2 \cdot H_2O$: C, 65.6; H, 4.7%.

When an aqueous solution of the sodium salt of the acid was heated, the alkaloid ($C_{14}H_8ON_2$, identified by m.p. and mixed m.p.) separated quantitatively. The alkaloid was also obtained after attempts to esterify the acid with methanolic hydrogen chloride and after heating a solution of the acid in acetic acid at its boiling point for 4 hours.

(ii) *trans- β -Carbolylacrylic Acid (IV)*. A solution of *cis*- β -carbolylacrylic acid in 20% sodium hydroxide was refluxed for 6 hours and then concentrated by heating on the water bath. The amorphous sodium salt which separated on cooling was filtered off, dissolved in water, and the solution acidified with acetic acid. The precipitated *trans*-acid was crystallized from aqueous acetic acid and again from water and obtained as flat yellow needles, m.p. 274 °C. (uncorr., decomp.). This melting point behaviour was unaltered by admixture with a specimen kindly supplied by Mr. G. K. Hughes.

Found: C, 68.0; H, 4.5; N, 11.3%.

Calculated for $C_{14}H_{10}O_2N_2 \cdot \frac{1}{2}H_2O$: C, 68.0; H, 4.5; N, 11.3%.

This acid was not reconverted to the alkaloid by heating a solution of its sodium salt in water, and it was esterified by methanolic hydrogen chloride. The *methyl ester* crystallized from aqueous methanol as yellow needles, m.p. 177–178 °C.

Found: C, 71.3; H, 4.9; CH_3O , 11.9%.

Calculated for $C_{15}H_{12}O_2N_2$: C, 71.4; H, 4.8; CH_3O , 12.3% (one methoxyl).

The ester dissolves readily in 0.01N hydrochloric acid.

(g) *Hydrogenation of the Alkaloid $C_{14}H_8ON_2$*

(i) The alkaloid (0.39 g.) in ethanol (40 ml.), hydrogenated in the presence of Raney nickel, absorbed 45.8 ml. H_2 at 21 °C. and 752 mm.; calculated for 1 mole, 43.2 ml. The solution was filtered, evaporated, and the residue crystallized from light petroleum giving *4,5-dihydrocanthin-6-one* (VIII) as colourless needles, m.p. 130–131 °C.

Found: C, 75.8; H, 4.5%.

Calculated for $C_{14}H_{10}ON_2$: C, 75.7; H, 4.5%.

Dihydrocanthinone is readily soluble in 0.01N hydrochloric acid.

(ii) The same compound was obtained in poor yield when the alkaloid (0.5 g.) in benzene (75 ml.) was refluxed with c. 4 g. Raney nickel for 3 hours. The nickel was washed several times with benzene by decantation and the combined benzene solutions evaporated leaving 0.18 g. colourless solid. After crystallization from benzene-light petroleum the substance melted at 130–131 °C. and did not depress the m.p. of the specimen prepared by hydrogenation in ethanol.

(iii) The alkaloid (1.5 g.) in acetic acid (15 ml.) was boiled with granulated zinc for 8 minutes, the solution diluted with water, neutralized by the addition of sodium bicarbonate, and extracted with chloroform. Evaporation of the chloroform and crystallization of the residue from light petroleum gave dihydrocanthinone, m.p. and mixed m.p. 129–130 °C.

(h) 2-(1'- β -Carbolyl)propionic Acid (IX)

(i) 4,5-Dihydrocanthinone dissolved in a little methanol was added to warm 5% sodium hydroxide and the methanol evaporated on the water-bath. The solution was acidified with acetic acid and the liberated acid crystallized from water. It was obtained as colourless needles, m.p. 215–216 °C. (decomp.). Analysis of the acid was unsatisfactory due to the retention of varying amounts of water. Consequently the *silver salt* was prepared and analysed.

Found: C, 47.9; H, 3.1; Ag, 31.1%.

Calculated for $C_{14}H_{11}O_2N_2Ag$: C, 48.4; H, 3.2; Ag, 31.1%.

The acid is very soluble in acetic acid, but insoluble in chloroform, benzene, or cold alcohol. Its aqueous solutions have a strong blue fluorescence. It does not dissolve in aqueous sodium bicarbonate, gives no colour with ferric chloride, and was recovered unchanged both after refluxing a solution in acetic acid for 4 hours and after refluxing an aqueous solution of the potassium salt for 4 hours. The acid was esterified with methanolic hydrogen chloride and the *methyl ester* obtained as colourless needles from benzene, m.p. 145–146 °C.

Found: C, 70.8; H, 5.8; N, 11.0; CH_3O , 12.2%.

Calculated for $C_{15}H_{14}O_2N_2$: C, 70.9; H, 5.5; N, 11.0; CH_3O , 12.2% (one methoxyl).

The *ethyl ester* was prepared by treating the acid with thionyl chloride followed by ethanol, and crystallized from benzene-light petroleum as colourless needles, m.p. 125–126 °C.

Found: C, 71.2; H, 5.9%.

Calculated for $C_{16}H_{16}O_2N_2$: C, 71.6; H, 6.0%.

The methyl ester was also prepared in this way.

(ii) *cis*- β -Carbolylacrylic acid (III, from 0.5 g. canthinone) dissolved in sodium hydroxide (8%, 25 ml.) was heated on the water-bath and Raney alloy (1.5 g.) added gradually during 3 hours. The cooled and filtered solution deposited a mass of colourless plates which were dissolved in water and the β -carbolylpropionic acid liberated with acetic acid. Yield 42% after crystallization from water. The identity was confirmed by m.p. and mixed m.p. of the methyl ester.

(iii) *trans*- β -Carbolylacrylic acid reduced as in (ii) likewise gave β -carbolylpropionic acid, the identity being confirmed by m.p. and mixed m.p. of the methyl ester.

(iv) The alkaloid (II, 0.1 g.) dissolved in methanol (50 ml.) was refluxed with Raney nickel (2 g.) for 4 hours. The product crystallized from benzene as colourless needles, m.p. 145–146 °C. undepressed by admixture with the methyl ester of β -carbolylpropionic acid.

(v) Canthinone hydrochloride dissolved in ethanol was hydrogenated using Adams's platinum catalyst and the product recrystallized from benzene-light petroleum. It separated as colourless needles, m.p. 126.5–127.5 °C., alone or in admixture with the ethyl ester of β -carbolylpropionic acid.

(vi) A solution of β -carbolylpropionic acid (0.9 g.) and selenium dioxide (0.9 g.) in water (25 ml.) was refluxed for 24 hours, filtered from selenium, and basified with sodium bicarbonate. The basic product was extracted with chloroform and a yellow acid, suspended in the aqueous layer, filtered off. Only a negligible additional quantity of acid resulted from acidification of the bicarbonate filtrate.

The base (0.26 g.) was purified by passage through the picrate and finally by crystallization from benzene-light petroleum. It melted at 161–162°C. undepressed by admixture with canthin-6-one.

Found: C, 76.6; H, 3.7%.

Calculated for $C_{14}H_8ON_2$: C, 76.4; H, 3.6%.

The identity was confirmed by the preparation of canthinone picrate, m.p. and mixed m.p. 262–264°C. (uncorr.).

The acid (0.64 g.) was crystallized from acetic acid and obtained as yellow needles, m.p. 273–274°C. (uncorr., decomp.). It was converted to its methyl ester, m.p. 177–178°C. undepressed by admixture with the methyl ester of *trans*-5-carbolylaerylic acid.

Found: C, 71.5; H, 4.7%.

Calculated for $C_{15}H_{12}O_2N_2$: C, 71.4; H, 4.8%.

(i) 4,5-Dihydrocanthine (XI)

A solution of canthinone (0.5 g.) in hydrochloric acid (5%, 50 ml.) was boiled with zinc dust for 15 minutes. The almost colourless solution was filtered and allowed to crystallize giving a pale yellow solid which was dissolved in water and the solution basified with sodium hydroxide. Extraction with chloroform yielded a sticky base which crystallized on standing in air (0.30 g.). After purification through the hydrochloride the base was crystallized from aqueous methanol giving dihydrocanthine hydrate as colourless plates, m.p. 74–75°C. (air-dried).

Found: C, 74.7; H, 6.0%.

Calculated for $C_{14}H_{12}N_2 \cdot H_2O$: C, 74.3; H, 6.2%.

When dried in a desiccator these crystals lost their form and became oily. *Dihydrocanthine picrate*, yellow needles from ethanol, melted at 258.5°C. (decomp.) after blackening at 254°C.

Found: C, 55.3; H, 3.4%.

Calculated for $C_{14}H_{12}N_2 \cdot C_6H_5O_7N_3$: C, 54.9; H, 3.4%.

Dihydrocanthine picrolonate, pale yellow needles from ethanol, m.p. 259–260°C. (decomp.).

Found: C, 61.2; H, 4.4%.

Calculated for $C_{14}H_{12}N_2 \cdot C_{10}H_8O_5N_4$: C, 61.0; H, 4.2%.

Dihydrocanthine is readily soluble in 0.01N hydrochloric acid. The *hydrochloride* separated from 0.5N acid as colourless needles, m.p. 282–283°C. (decomp., uncorr.).

Found: Cl, 11.9%.

Calculated for $C_{14}H_{12}N_2 \cdot HCl \cdot 3H_2O$: Cl, 11.9%.

Its aqueous solutions have an intense blue fluorescence. Dihydrocanthine was recovered unchanged (m.p. and mixed m.p. of the picrate) after refluxing for 2 hours with 10% aqueous sodium hydroxide containing sufficient methanol to keep the base in solution. No acidic product was detected.

(i) Dihydrocanthine (0.4 g.) was refluxed for 4 hours with benzaldehyde (freshly distilled; 2.5 ml.), the mixture diluted with ether and shaken with 5% hydrochloric acid. A yellow precipitate of hydrochloride which formed at once was separated and washed with ether. Regeneration of the base gave a yellowish brown oil which was dissolved in benzene and chromatographed on alumina. The main fraction crystallized from benzene-light petroleum giving *benzaldihydrocanthine* as pale yellow plates, m.p. 140.5–141°C.

Found: C, 85.2; H, 5.6; N, 9.4%.

Calculated for $C_{21}H_{16}N_2$: C, 85.1; H, 5.4; N, 9.5%.

(j) Canthinone-N-oxide (X)

Hydrogen peroxide (30%; 0.35 ml.) was added to a solution of canthinone (0.2 g.) in acetic acid (5 ml.) and after 8 days at room temperature the solvent was evaporated under reduced pressure, the residue triturated with aqueous sodium bicarbonate, and then extracted with

chloroform. Evaporation of the chloroform left 0.22 g. solid which crystallized from benzene as yellow needles, m.p. 237.5–238.5 °C. (decomp.).

Found: C, 71.5; H, 3.7; N, 11.6%.

Calculated for $C_{14}H_8O_2N_2$: C, 71.2; H, 3.4; N, 11.9%.

Canthinone-N-oxide dissolved slowly in hot 5% sodium hydroxide to a yellow solution from which it was recovered unchanged after acidification. It was also recovered unchanged after a solution in methanol saturated with sulphur dioxide was allowed to stand for 3 days at room temperature. A solution in acetic acid was boiled for 5 minutes with zinc dust, poured into water, and basified with sodium bicarbonate. Extraction with chloroform and crystallization of the extracted base from light petroleum gave dihydrocanthinone, m.p. and mixed m.p. 128–130 °C.

(k) *Attempted Dehydrogenation of Hexahydrocanthinone*

(i) Hexahydrocanthinone (53 mg.), prepared as described by Hahn and Hansel (1938), and chloranil (116 mg.) were dissolved in xylene (10 ml.) and the solution boiled under reflux for 2 hours. The black precipitate which formed rapidly was filtered, washed, and extracted twice with 1% hydrochloric acid. The acid extracts were basified, shaken with chloroform, and the recovered base converted to its hydrochloride, m.p. 280–281 °C. (uncorr.) alone or mixed with the hydrochloride of hexahydrocanthinone. No basic material was recovered from the xylene filtrate.

(ii) Following the procedure of Akabori and Saito (1930), hexahydrocanthinone (50 mg.) dissolved in a solution of maleic acid (78 mg.) in water (2 ml.) was refluxed for 5 hours with 0.1 g. 5% palladium-norit catalyst. The recovered base (24 mg.) melted at 148.5–149 °C. undepressed by admixture with the starting material. The same result was obtained when palladium black was employed as catalyst.

(iii) The procedure used by Hoshino and Takiura (1936) for dehydrogenating tetrahydrocarbazoles was applied as follows: hexahydrocanthinone (45 mg.), cinnamic acid (130 mg.), and a small quantity of platinum black were well mixed and heated from 140–180 °C. over a period of 10 minutes, during which time the mixture became dark green. It was then dissolved in benzene and the solution extracted with dilute hydrochloric acid. The small amount of basic material recovered was identified as hexahydrocanthinone by conversion to the hydrochloride (m.p. and mixed m.p.). The activity of the palladium black used in this and the preceding experiment was checked by using it to dehydrogenate tetrahydrocarbazole.

(iv) Potassium dichromate solution (10%; 0.9 ml.) was added dropwise during 5 minutes to a boiling solution of hexahydrocanthinone (48 mg.) in water (10 ml.) and acetic acid (0.5 ml.). Boiling was continued for a further 2 minutes but no change was observed and the base was recovered unchanged. Dehydrogenation likewise failed to occur when 1-(ω -carboxyethyl)-1,2,3,4-tetrahydro- β -carboline-1-carboxylic acid (XII) was boiled with aqueous potassium dichromate.

(l) *Ultraviolet Absorption Spectra*

The ultraviolet absorption spectra were measured in dioxan solution by means of a Beckmann model DU spectrophotometer.

IV. ACKNOWLEDGMENTS

The authors are indebted to Mr. L. J. Webb, Division of Plant Industry, C.S.I.R.O., for obtaining supplies of the necessary plant materials; to Professor E. J. Hartung for providing accommodation and facilities in the Chemistry Department, University of Melbourne; to Mr. J. Shelton, Division of Industrial Chemistry, C.S.I.R.O., for measuring the ultraviolet absorption spectrum of dihydrocanthinone; and to Associate Professor F. H. Shaw for carrying out pharmacological tests.

V. REFERENCES

- AKABORI, S., and SAITO, K. (1930).—*Ber. dtsh. chem. Ges.* **63**: 2245.
- ENGLER, A., and PRANTL, K. (1931). "Die Natürlichen Pflanzenfamilien." Vol. 19a. (Wilhelm Engelmann: Leipzig.)
- HAHN, G., and HANSEL, A. (1938).—*Ber. dtsh. chem. Ges.* **71**: 2163.
- HOSHINO, T., and TAKIURA, K. (1936).—*Bull. Chem. Soc. Japan* **11**: 218.
- KERMACK, W. O., PERKIN, W. H., and ROBINSON, R. (1921).—*J. Chem. Soc.* **119**: 1602.
- MANSKE, R. H., and MARION, L. (1938).—*Canad. J. Res. B* **16**: 432.
- OCHIAI, I., ISHIKAWA, M., and KATADA, M. (1943).—*J. Pharm. Soc. Japan* **63**: 307. (*Chem. Abstr.* **45**: 5153 (1951).)
- PATTERSON, A. M., and CAPELL, L. T. (1940).—"The Ring Index." (Reinhold Publ. Corp.: New York.) *
- PRELOG, V., SZPILFOGEL, S., and BATTEGAY, J. (1947).—*Helv. Chim. Acta* **30**: 366.
- SNYDER, H. R., WALKER, H. G., and WERBER, F. X. (1949).—*J. Amer. Chem. Soc.* **71**: 527.

ALKALOIDS OF THE AUSTRALIAN RUTACEAE :
EVODIA XANTHOXYLOIDES F. MUELL.

II. ISOLATION OF THE ALKALOIDS FROM THE LEAVES

By G. K. HUGHES,* K. G. NEILL,* and E. RITCHIE*

[Manuscript received December 17, 1951]

Summary

From the leaves of *Evodia xanthoxyloides* nine alkaloids, evoxanthine, melicopidine, evoxine, xanthevodine, xanthoxoline, evoxanthidine, evodine, evolidine, and evoxoidine were isolated. There was also obtained a tenth nitrogenous substance (X) which was shown to be 4-hydroxy-2,3-dimethoxy-10-methylacridone. This is not claimed to be an alkaloid as it may have been formed by demethylation during the isolation.

I. INTRODUCTION

In Part I of this series (Hughes and Neill 1949)† the isolation from the bark of *Evodia xanthoxyloides* F. Muell. of the alkaloids, evoxanthine, melicopidine, and kokusaginine was reported. A detailed examination of the leaves showed the presence of nine alkaloids and a further substance which may be an alkaloid or a demethylation product of one. Evoxanthine and melicopidine were found but kokusaginine was not detected.

The new alkaloids have been named *evoxine* $C_{16}H_{21}O_6N$, *xanthevodine* $C_{16}H_{13}O_5N$, *evoxanthidine* $C_{15}H_{11}O_4N$, *xanthoxoline* $C_{15}H_{13}O_4N$, *evodine* $C_{18}H_{19}O_5N$ (?), *evoxoidine* $C_{15}H_{15}O_4N$ (?), and *evolidine* $C_{15}H_{23}O_4N$.

The alkaloid content was approximately 2 per cent. by weight of the dried leaves, the major constituents being evoxine (0.9 per cent.) and evoxanthine (0.6 per cent.).

The methanolic extract of the leaves was concentrated as far as possible and the residue boiled with 5 per cent. hydrochloric acid. Although it was realized that this was undesirable owing to the easy demethylation of the acridone alkaloids, it proved to be the only practicable method of obtaining the alkaloids free from chlorophyll and much resinous material. After basifying the acid extract the residue was refluxed with much chloroform and filtered from tar.

The separation of the alkaloids was achieved by empirical methods based primarily on their extraction from either chloroform or benzene solution by various strengths of hydrochloric acid. Use was also made of the easy demethylation of evoxanthine and melicopidine to their nor-compounds; these only dissolved in concentrated acid. The last traces of nor-compounds could also be

* Department of Organic Chemistry, University of Sydney.

† HUGHES, G. K., and NEILL, K. G. (1949).—*Aust. J. Sci. Res. A* 2: 429.

conveniently removed by adsorption on alumina. This was the only satisfactory method for the purification of xanthevodine.

As may be expected the separation was both complex and laborious: it was often found that small amounts of some of the alkaloids occurred in fractions not expected from subsequent knowledge of their properties. This was particularly the case with xanthevodine which was often most tedious to remove. It

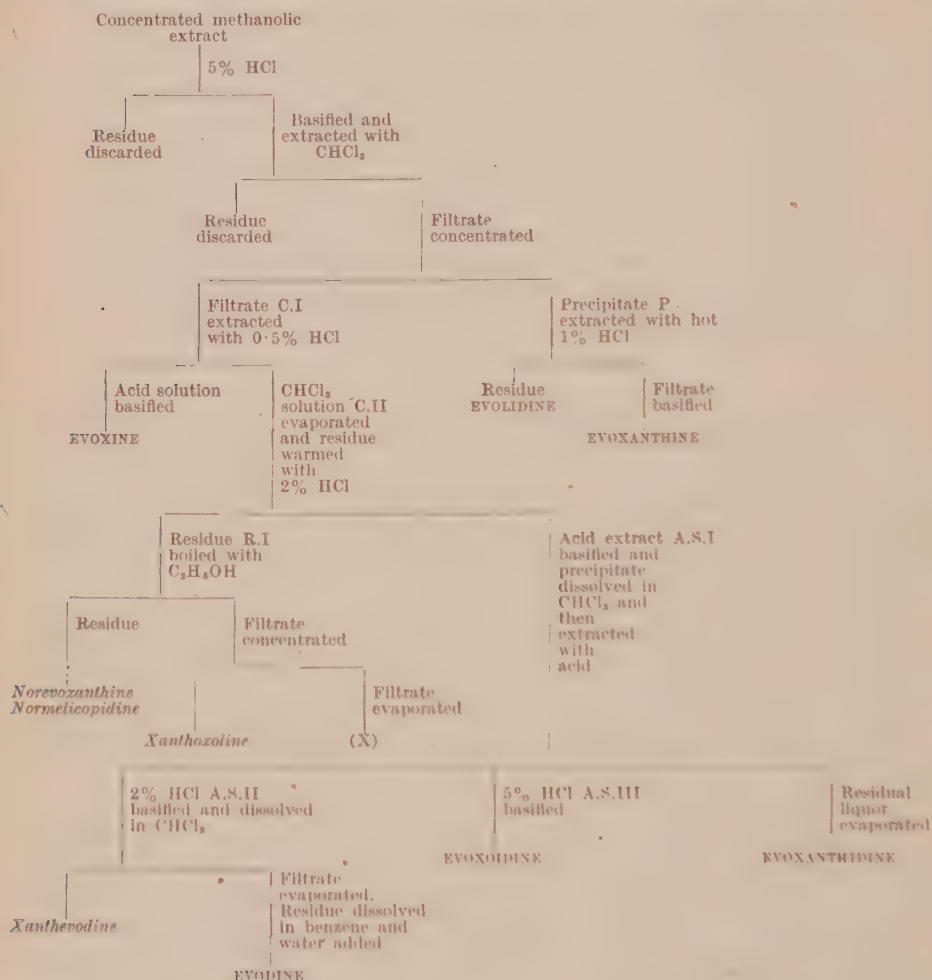


Fig. 1

should also be understood that the methods outlined in Section II and the chart (Fig. 1) were repeated several times in order to achieve the final result.

It was early recognized that xanthoxoline had properties similar to the noralkaloids and thus it may have been an artefact. However, by using warm acetic acid instead of boiling 5 per cent. hydrochloric acid in the original extraction it was possible to obtain some xanthoxoline. This is regarded as sufficient evidence to claim that it occurs as such in the plant. All efforts to

isolate the tenth compound (X) by this method failed. An attempt to separate the original crude mixture of bases by adsorption on alumina proved abortive.

Compound X, $C_{17}H_{17}O_4N$ (yield 0.03 per cent.), crystallized from ethanol in golden needles, m.p. 176–177 °C. This was identified as 4-hydroxy-2,3-dimethoxy-10-methylacridone previously prepared from evoxanthine. Further proof of its structure was obtained by its oxidation to 2-methoxy-10-methylacridone-3,4-quinone and its methylation to 2,3,4-trimethoxy-10-methylacridone. This nor-compound has also been isolated from the bark by alcoholic extraction of the nor-fractions.

II. EXPERIMENTAL

All melting points are uncorrected. Analyses are by Dr. Weiler and Dr. Strauss, Oxford.

The dried, milled leaves, which also contained some seeds, (13 kg.) were exhausted by cold percolation with methanol. After concentration the residue was repeatedly extracted with hot 5% hydrochloric acid (12 l.). The acid extract after standing in the refrigerator was decanted from the black tar and basified. The aqueous layer was separated from the dark viscous precipitate and thoroughly extracted with chloroform. The precipitate was then refluxed with the combined chloroform extracts and the mixture filtered. The filtrate was concentrated to about 3 l. and allowed to stand. The crystalline precipitate (P) was collected leaving the filtrate (C.I.).

(a) Precipitate (P)

This was treated three times with hot 1% hydrochloric acid and the residue collected. The residue was boiled with acetone which dissolved most of the colour leaving an off-white crystalline mass. This was recrystallized from methanol (charcoal), and evolidine obtained as very fine white, matted needles, m.p. 287–288 °C. Yield 0.05%.

Found: C, 58.9; H, 7.5; N, 13.1%; CH_3O , nil; CH_3N , nil.

Calculated for $C_{15}H_{23}O_4N_3$: C, 58.7; H, 7.5; N, 13.7%.

Evolidine is a very weak base soluble only in concentrated acid. Attempts were made to obtain further amounts of it from the original residues but these were unsuccessful. A further quantity was obtained as described in Section II (c).

When pure it is almost insoluble in chloroform and acetone and its alcoholic solutions show no fluorescence. The 1% hydrochloric acid extracts were basified with ammonia. The precipitate consisted almost entirely of evoxanthine plus some nor-compounds. Evoxanthine was purified either from ethanol or chloroform as yellow needles, m.p. and mixed m.p. 217–218 °C. with an authentic specimen from the bark.

(b) Chloroform Solution (C.I.)

This was extracted with 0.5% hydrochloric acid until only a faint precipitate was formed on making alkaline. The combined acid extracts were basified with ammonia and the faintly yellow crystalline precipitate collected. The roughly dried solid was then repeatedly extracted with boiling benzene. On cooling, evoxine crystallized in colourless needles which, after recrystallization (charcoal) from benzene, ethanol, or preferably ethyl acetate, had m.p. 154–155 °C.

Found: C, 58.9; H, 6.5; N, 4.5; CH_3O , 19.2%.

Calculated for $C_{16}H_{21}O_6N$: C, 59.4; H, 6.5; N, 4.3; $2 \times CH_3O$, 19.2%.

$[\alpha]_D^{22} + 14.6^\circ$, $[\alpha]_{5461}^{22} + 42.7^\circ$ (c 0.48 in alcohol).

Evoxine was easily soluble in cold chloroform. It gave a negative Labat test for a methylenedioxy-group and a cherry red colour with concentrated nitric acid. An ethanolic solution showed an intense blue-violet fluorescence.

The benzene filtrate was concentrated and yielded some impure evoxine mixed with a little xanthevodine, evoxanthine, and nor-compounds. All these were preferentially adsorbed on alumina before the evoxine was recrystallized.

(c) *Chloroform Solution (C.II)*

This was evaporated to dryness and the residue thoroughly extracted with warm 2% hydrochloric acid (A.S.I). The yellow residue (R.I) was refluxed with much ethanol and filtered. The residue (R.II) was essentially norevioxanthine and normelicopidine with a small amount of evolidine. The filtrate on concentration and cooling gave crude xantholine. The mother liquors from this were evaporated to dryness and the bright yellow residue boiled with acetone and filtered. Compound X was precipitated on cooling and after several recrystallizations from ethanol was obtained as long golden needles, m.p. and mixed m.p. 176–177 °C. with an authentic specimen of 4-hydroxy-2,3-dimethoxy-10-methylacridone. Yield 0.03%.

Found: C, 66.8; H, 5.5; N, 5.1; CH_3O , 21.8%.

Calculated for $\text{C}_{16}\text{H}_{16}\text{O}_4\text{N}$: C, 67.3; H, 5.3; N, 4.9; $2 \times \text{CH}_3\text{O}$, 21.8%.

Compound X was methylated by refluxing an acetone solution with methyl sulphate and potassium carbonate. After recrystallization from benzene-light petroleum the product was obtained as pale yellow plates, m.p. and mixed m.p. 168.5–170 °C. with an authentic specimen of 2,3,4-trimethoxy-10-methylacridone prepared from evoxanthine (*loc. cit.*).

Compound X (0.2 g.) was oxidized with cold concentrated nitric acid (5 cc.) and the product was recrystallized from water as dark red needles, m.p. and mixed m.p. 279–280 °C. with an authentic specimen of 2-methoxy-10-methylacridone-3,4-quinone.

The crude xanthoxoline was recrystallized first from glacial acetic acid and then from much ethanol as golden needles which had a double melting point, firstly 256–257 °C. and then, if allowed to solidify, at 265–267 °C. Yield 0.1%.

Found: C, 66.4; H, 4.7; N, 5.0; CH_3O , 21.8%; CH_3N , nil.

Calculated for $\text{C}_{15}\text{H}_{13}\text{O}_4\text{N}$: C, 66.4; H, 4.8; N, 5.2; CH_3O , 22.8%.

Xanthoxoline was soluble in hot alcohol, sparingly soluble in chloroform, and almost insoluble in benzene. It gave a negative test for a methylenedioxy-group. An ethanolic solution showed no fluorescence but gave an intense dark green colour with ferric chloride. It is a weak base dissolving only in concentrated acid.

(d) *Acid Solution (A.S.I)*

This was basified with ammonia and the precipitate transferred to chloroform. The chloroform solution was extracted several times with 2% hydrochloric acid (A.S.II) then 5% hydrochloric acid (A.S.III). During these extractions a substance appeared at the interface. This was later shown to be evoxanthidine which was the main constituent of the residual chloroform solution. Evoxanthidine crystallized from alcohol in long pale yellow needles, m.p. 312–313 °C. Yield 0.03%.

Found: C, 66.5; H, 4.1; N, 5.1; CH_3O , 11.6%; CH_3N , nil.

Calculated for $\text{C}_{16}\text{H}_{11}\text{O}_4\text{N}$: C, 66.9; H, 4.1; N, 5.2; $1 \times \text{CH}_3\text{O}$, 11.5%.

It was only slightly soluble in chloroform and less so in benzene. It gave a positive test for a methylenedioxy-group and a dilute solution in ethanol showed an intense blue fluorescence. It gave no colour reaction with ferric chloride.

(e) *Acid Solution (A.S.II)*

This was basified, the precipitate dried and then dissolved in chloroform. On concentration and cooling crude xanthevodine was obtained. This was recrystallized from chloroform and then ethanol giving large bright yellow prisms, m.p. 213–214 °C. Yield 0.2%.

Found: C, 64.0; H, 4.3; N, 4.8; CH_3O , 20.4%; CH_3N , nil.

Calculated for $\text{C}_{16}\text{H}_{13}\text{O}_5\text{N}$: C, 64.2; H, 4.4; N, 4.7; $2 \times \text{CH}_3\text{O}$, 20.7%.

Xanthevodine was previously assigned (Hughes and Neill 1949) the formula $\text{C}_{17}\text{H}_{15}\text{O}_5\text{N}$ but further analyses are in better agreement with the formula $\text{C}_{16}\text{H}_{13}\text{O}_5\text{N}$. In dilute ethanolic solution it showed a yellow-green fluorescence. It formed a hydrochloride, m.p. 137 °C. (decomp.), which was sparingly soluble in water and a picrate, m.p. 145–146 °C. It gave a positive test for a methylenedioxy-group.

The chloroform extracts were evaporated to dryness and the residue taken up in benzene. This was passed through a short column of alumina which removed some of the colour from the solution. On shaking the eluate with water an off-white crystalline substance separated. This was collected and recrystallized from methanol (charcoal) as colourless prisms, m.p. 153–154 °C. Yield 0.01%.

Evodine was soluble in chloroform and dry benzene and gave a negative methylenedioxy test.

Found: C, 65.7; H, 6.0; N, 4.2; CH_3O , 16.4%.

Calculated for $\text{C}_{18}\text{H}_{18}\text{O}_5\text{N}$: C, 65.6; H, 5.8; N, 4.3; $2 \times \text{CH}_3\text{O}$, 18.9%.

(f) *Acid Solution (A.S.III)*

The solution was made alkaline with ammonia and the precipitate collected and dried. It was then dissolved in benzene and cooled when a further small quantity of xanthevodine was obtained. The filtrate was extracted with a small volume of 2% hydrochloric acid to remove traces of evoxanthine and then with 5% hydrochloric acid. The latter solution was then exhausted with chloroform. This solution was passed through a column of alumina and the almost colourless eluate then evaporated to dryness. The residue on recrystallization from methanol gave evoxoidine as stout colourless needles, m.p. 136–137 °C. Yield 0.01%.

Found: C, 65.2; H, 5.7; N, 5.2; CH_3O , 17.5%; CH_3N , nil.

Calculated for $\text{C}_{18}\text{H}_{18}\text{O}_4\text{N}$: C, 65.9; H, 5.6; N, 5.1; $2 \times \text{CH}_3\text{O}$, 22.7%.

Evoxoidine was easily soluble in chloroform and benzene and gave a negative test for a methylenedioxy-group.

III. ACKNOWLEDGMENTS

The authors wish to thank Mr. L. J. Webb, C.S.I.R.O., and the Queensland Department of Forestry for collecting the leaves used in this examination.

ALKALOIDS OF THE AUSTRALIAN RUTACEAE:
EVODIA XANTHOXYLOIDES F. MÜELL.

III. THE STRUCTURES OF THE COLOURED ALKALOIDS, EVOXANTHIDINE,
XANTHEVODINE, AND XANTHOXOLINE

By J. R. CANNON,* G. K. HUGHES,* K. G. NEILL,* and E. RITCHIE*

[Manuscript received December 17, 1951]

Summary

The coloured alkaloids, evoxanthidine, xanthevodine, and xanthoxoline, were shown to be 4-methoxy-2,3-methylenedioxyacridone, 1,4-dimethoxy-2,3-methylenedioxyacridone, and 2,3-dimethoxy-4-hydroxyacridone, respectively. A total synthesis of xanthoxoline was achieved.

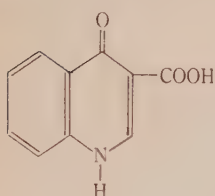
I. INTRODUCTION

The isolation of the coloured alkaloids, evoxanthidine $C_{15}H_{11}O_4N$, xanthevodine $C_{16}H_{13}O_5N$, and xanthoxoline $C_{15}H_{13}O_4N$, from the leaves of *Erodia xanthoxyloides* F. Muell. was described in Part II of this series (Hughes, Neill, and Ritchie 1952). Evoxanthidine forms pale yellow needles and has an intense blue fluorescence in dilute alcoholic solution. Analysis revealed the presence of one methoxyl group, and the absence of a methylimino-group. The Labat methylenedioxy test (gallic acid-sulphuric acid) was positive. Xanthevodine is also pale yellow and its dilute alcoholic solution exhibits a yellowish green fluorescence. Two methoxyls and a methylenedioxy-group are present, but no methylimino-group was found. Xanthoxoline crystallizes in golden-yellow needles which show no fluorescence in alcoholic solution, in contrast to the above two alkaloids. Two methoxyl groups were found, but neither methylimino- nor methylenedioxy group was present. It is a much weaker base than either evoxanthidine or xanthevodine, being insoluble in hot 10 per cent. hydrochloric acid. The presence of a masked phenolic group was suspected, for although the alkaloid was practically insoluble in 10 per cent. aqueous sodium hydroxide, an alcoholic solution gave an intense green coloration with aqueous ferric chloride. Also, on acetylation, xanthoxoline yielded an acetate, $C_{17}H_{15}O_5N$, the alcoholic solution of which had a blue fluorescence and gave no coloration with ferric chloride. Xanthoxoline was regenerated from the acetate by alkaline hydrolysis.

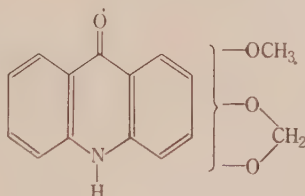
In their general properties evoxanthidine and xanthevodine resemble the 10-methylacridone alkaloids, whereas the behaviour of xanthoxoline is reminiscent of the nor-alkaloids of this series (Crow and Price 1949; Hughes and Neill 1949). All three alkaloids were submitted to nitric oxide oxidation

* Department of Organic Chemistry, University of Sydney.

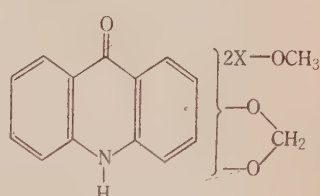
when both evoxanthidine and xanthevodine were converted to an acid, $C_{10}H_7O_3N$, which was identified as 4-quinolone-3-carboxylic acid (I). Hence evoxanthidine (II) and xanthevodine (III) are acridone derivatives substituted in only one ring of the nucleus.



(I)



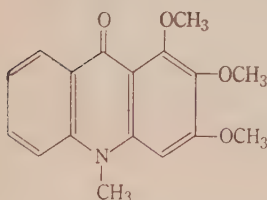
(II)



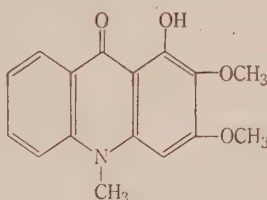
(III)

Xanthoxoline yielded a red amorphous solid which could not be crystallized and was not investigated further.

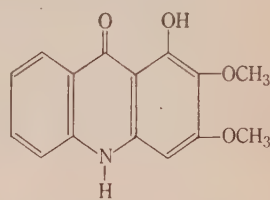
An attempt to methylate the weakly phenolic group of xanthoxoline by refluxing with methyl iodide and potassium carbonate in acetone produced, in good yield, a mixture of two substances which were easily separated by virtue of their widely different basicities. The stronger base, which was soluble in warm 3N acetic acid was identified as 2,3,4-trimethoxy-10-methylacridone (IV)



(IV)

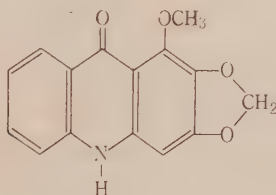


(V)

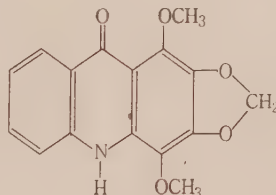


(VI)

and the less basic was found to be 2,3-dimethoxy-4-hydroxy-10-methylacridone (V). N-methylation must have occurred before O-methylation in this case, as V yields IV on further methylation, and thus xanthoxoline must be 2,3-dimethoxy-4-hydroxyacridone (VI).



(VII)



(VIII)

The various methods available for the N-methylation of acridones have been reviewed by Hughes *et al.* (1952) but methyl iodide and potassium carbonate in acetone had not been used previously. By means of this method evoxanthidine and xanthevodine were converted to evoxanthine and melicopidine, respectively,

in good yield. Hence evoxanthidine is 4-methoxy-2,3-methylenedioxyacridone (VII) and xanthevodine is 1,4-dimethoxy-2,3-methylenedioxy-acridone (VIII).

Although both evoxanthidine and xanthevodine contain a methoxyl in the 4-position, only partial demethylation occurred when the alkaloids were refluxed with alcoholic hydrochloric acid for 8 hours, in contrast to the corresponding 10-methylacridones which are almost quantitatively demethylated after 1 hour (Crow and Price 1949; Hughes and Neill 1949). However, when the hydrochlorides were heated at 170–180°C. for a short time, good yields of the corresponding nor-compounds were obtained. *Nor-evoxanthidine* $C_{14}H_9O_4N$ and *norxantherodine* $C_{15}H_{11}O_5N$ each had the characteristic properties of the nor-10-methylacridones, namely, a higher melting point and deeper colour than the parent base, low solubility in the common organic solvents and very weak basicity. Also, the ethanolic solutions did not fluoresce and gave dark green colorations with ferric chloride.

Confirmation of the structure (VI) of xanthoxoline was obtained by the following synthesis. 3,4,5-Trimethoxydiphenylamine-2'-carboxylic acid on treatment with phosphorus oxychloride underwent ring closure to 2,3,4-trimethoxy-5-chloroacridine. Hydrolysis with concentrated hydrochloric acid after the method previously described (Hughes, Neill, and Ritchie 1950) yielded a mixture containing xanthoxoline as well as the expected 2,3,4-trimethoxyacridone which was eventually obtained in good yield by refluxing the 2,3,4-trimethoxy-5-chloroacridine with 1% hydrochloric acid for half an hour. When the dry hydrochloride of 2,3,4-trimethoxyacridone was heated at 150–160°C. for a short time, demethylation occurred, yielding 2,3-dimethoxy-4-hydroxyacridone which proved to be identical with natural xanthoxoline.

II. EXPERIMENTAL

Melting points are uncorrected. Analyses are by Dr. Weiler and Dr. Strauss, Oxford.

(a) 4-Quinolone-3-carboxylic Acid

To either evoxanthidine or xanthevodine (1 g.) was added concentrated nitric acid (10 ml.); the mixture reacted vigorously and the excess nitric acid was boiled off as quickly as possible. The residual brown gum crystallized from glacial acetic acid and, after recrystallization from water, colourless needles, m.p. 266°C. (decomp.) were obtained. (Camps (1901) found m.p. 266°C.).

Found: C, 63.8; H, 3.7%.

Calculated for $C_{10}H_7O_3N$: C, 63.5; H, 3.7%.

(b) Acetylation of Xanthoxoline

Xanthoxoline (1 g.), acetic anhydride (30 ml.), and fused sodium acetate (3 g.) were refluxed for 8 hours. Isolation of the product by pouring the reaction mixture into water, followed by recrystallization from ethyl acetate gave pale yellow prisms, m.p. 238–241°C., in good yield.

Found: C, 65.1; H, 4.7; N, 4.8; CH_3O , 19.6%.

Calculated for $C_{17}H_{16}O_5N$: C, 65.2; H, 4.8; N, 4.5; $2 \times CH_3O$, 19.8%.

A dilute ethanolic solution of xanthoxoline acetate gave no coloration with ferric chloride and showed a weak purple fluorescence which became intensely yellowish green on adding a drop of 10% sodium hydroxide. Hydrolysis of xanthoxoline acetate (0.5 g.) with 10% methanolic potassium hydroxide (25 ml.) yielded xanthoxoline, m.p. and mixed m.p. 266–268°C. (decomp.).

(c) *Methylation of Xanthoxoline*

Xanthoxoline (1.5 g.) in acetone (200 ml.) was refluxed with methyl iodide (30 ml.) and anhydrous potassium carbonate (40 g.) for 24 hours. The mixture was filtered, the salts washed with acetone, the yellow filtrate evaporated on the water-bath, and the residue extracted with warm 3N acetic acid (100 ml. in 3 portions). The insoluble residue was crystallized from ethanol when long, golden needles (0.48 g.) were obtained, m.p. and mixed m.p. with authentic 2,3-dimethoxy-4-hydroxy-10-methylacridone (Hughes and Neill 1949) 176–177 °C. The acetic acid solution was basified with ammonia, the precipitated base dissolved in benzene, the solution passed through a column of alumina, and the chromatogram developed with chloroform. Evaporation of the eluate and recrystallization of the residue from benzene-light petroleum yielded very pale yellow plates (0.67 g.), m.p. and mixed m.p. with 2,3,4-trimethoxy-10-methylacridone (Hughes and Neill 1949) 168–170 °C.

(d) *Methylation of 2,3-Dimethoxy-4-hydroxy-10-methylacridone*

On refluxing 2,3-dimethoxy-4-hydroxy-10-methylacridone (1 g.) with methyl iodide (30 ml.), acetone (200 ml.), and anhydrous potassium carbonate (40 g.) for 30 hours, and isolating the fully methylated compound by chromatography on alumina as in Section II (c), 2,3,4-trimethoxy-10-methylacridone (0.55 g.) was obtained, m.p. and mixed m.p. 168–170 °C.

(e) *Methylation of Evoxanthidine*

Evoxanthidine (1 g.) was suspended in acetone (200 ml.), methyl iodide (30 ml.) and anhydrous potassium carbonate (40 g.) were added, and the mixture refluxed for 33 hours. The salts were filtered off and washed with boiling chloroform, the filtrate evaporated, the residue extracted with boiling chloroform, the extract evaporated, and the residue crystallized from ethanol (15 ml.). Pale yellow needles (0.67 g.) were obtained, m.p. and mixed m.p. with evoxanthine 217–218 °C.

The identity of the methylated substance was confirmed by conversion to norevioxanthine (Hughes and Neill 1949), m.p. and mixed m.p. 274–276 °C.

(f) *Methylation of Xanthevodine*

Xanthevodine (1 g.) in acetone (150 ml.) was refluxed with methyl iodide (30 ml.) and anhydrous potassium carbonate (40 g.) for 33 hours. The filtrate from the insoluble salts was evaporated and the residue extracted with boiling benzene. Concentration of the benzene solution followed by the addition of light petroleum gave yellow prisms (0.81 g.), m.p. and mixed m.p. with a specimen of natural melicopidine 118–120 °C.

Conversion to normelicopidine in the usual manner (Crow and Price 1949) gave orange needles, m.p. and mixed m.p. 208–209 °C.

When methyl iodide was replaced by dimethyl sulphate in the above, melicopidine was obtained in lower yield (57%) and was accompanied by dark green coloured by-products which were difficult to remove.

The N-methylation of evoxanthidine and xanthevodine was also effected by heating the dry potassium salt of the acridone with dimethyl sulphate, but the yield was very poor in each case.

(g) *Norevioxanthidine*

Evoxanthidine (1 g.) was converted to the hydrochloride by warming with concentrated hydrochloric acid. The salt was then heated at 170–180 °C. for half an hour and the product crystallized from aqueous dioxane. Golden needles, m.p. 327 °C. (decomp.) were obtained.

Found: C, 66.4; H, 3.6; N, 5.8%.

Calculated for $C_{14}H_9O_4N$: C, 65.9; H, 3.6; N, 5.5%.

The dilute alcoholic solution did not fluoresce but gave a deep green colour with ferric chloride.

(h) *Norxanthevodine*

Xanthevodine hydrochloride (prepared by dissolving xanthevodine (1 g.) in hot 2% hydrochloric acid) was heated at 170–180 °C. for 5 minutes when it melted with effervescence, then resolidified. The product was extracted with warm 2% hydrochloric acid and the residue crystallized from aqueous dioxane, forming orange needles, m.p. 272–273 °C.

Found: N, 4.6; CH₃O, 10.8%.

Calculated for C₁₅H₁₁O₅N: N, 4.9; 1 × CH₃O, 10.9%.

(i) *Acetyl-norxanthevodine*

Norxanthevodine was acetylated as in Section II (b). The product crystallized from ethyl acetate as golden-yellow plates, m.p. 267–269 °C.

Found: C, 62.5; H, 4.2; N, 4.4%.

Calculated for C₁₇H₁₃O₆N: C, 62.4; H, 4.0; N, 4.3%.

The dilute ethanolic solution gave no colour with ferric chloride. When one drop of 10% aqueous sodium hydroxide was added to the cold ethanolic solution a brilliant blue fluorescence was noted.

(j) *Methylation of Norxanthevodine*

Norxanthevodine (0.5 g.) in acetone (150 ml.) was refluxed with methyl iodide (35 ml.) and anhydrous potassium carbonate (20 g.) for 30 hours. The mixture was worked up as in Section II (f) to yield melicopidine (0.34 g.), m.p. and mixed m.p. 118–119 °C.

(k) *2,3,4-Trimethoxy-5-chloroacridine*

3,4,5-Trimethoxydiphenylamine-2'-carboxylic acid (12 g.) (Hughes, Neill, and Ritchie 1950) was refluxed with phosphorus oxychloride (150 ml.) in an oil bath at 140 °C. for 3 hours, the excess phosphorus oxychloride removed on the water-bath under reduced pressure, and the red syrupy residue poured into a mixture of ice (500 g.), ammonia (600 ml. of d. 0.880), and chloroform (500 ml.), with vigorous stirring. The chloroform layer was separated, dried with sodium sulphate, evaporated, and the residue crystallized from light petroleum when 2,3,4-trimethoxy-5-chloroacridine was obtained as pale yellow, matted needles (9.5 g.), m.p. 146–147 °C.

Found: C, 63.3; H, 4.8%.

Calculated for C₁₆H₁₁O₃NCl: C, 63.3; H, 4.6%.

(l) *2,3,4-Trimethoxyacridone*

2,3,4-Trimethoxy-5-chloroacridine (5 g.) was dissolved in 1% hydrochloric acid (200 ml.) and the scarlet solution refluxed for half an hour. The solution, which was now orange in colour, was cooled, basified with ammonia, and the precipitated base extracted with chloroform. The solution was dried over sodium sulphate and run through a column of alumina. A bright yellow, practically stationary band developed at the top of the column (xanthoxoline?). Evaporation of the eluate to 30 ml. followed by the addition of benzene (100 ml.) yielded 2,3,4-trimethoxyacridone (4.5 g.) as very pale yellow needles, m.p. 208–210 °C. (Hughes, Neill, and Ritchie 1950 found m.p. 208–209 °C.).

The ethanolic solution showed a purple fluorescence and gave no coloration with ferric chloride.

(m) *2,3-Dimethoxy-4-hydroxyacridone (Xanthoxoline)*

2,3,4-Trimethoxyacridone (4 g.) was converted to the hydrochloride by adding concentrated hydrochloric acid (3 ml.) to a warm acetone solution (80 ml.). The yellow crystalline salt was washed with acetone then heated at 150–160 °C. for a short time until all frothing had ceased and the product had resolidified. After recrystallizations from ethanol and acetic acid, 2,3-dimethoxy-4-hydroxyacridone (3.1 g.) was obtained as golden-yellow flat needles, m.p. and mixed m.p. with natural xanthoxoline 265–268 °C. (decomp.) with slight previous softening.

The crude synthetic xanthoxoline showed the same curious double m.p. (254–256 °C., then melt crystallized and remelted at 265–267 °C. (decomp.)) as did crude, natural xanthoxoline. The ethanolic solution of the synthetic alkaloid was not fluorescent and gave a deep green colour with ferric chloride.

Found: C, 66.3; H, 5.1; N, 5.0; CH_3O , 21.6%.

Calculated for $\text{C}_{15}\text{H}_{13}\text{O}_4\text{N}$: C, 66.4; H, 4.8; N, 5.2; $2 \times \text{CH}_3\text{O}$, 22.8%.

(n) *Acetylation of Synthetic 2,3-Dimethoxy-4-hydroxyacridone*

Synthetic xanthoxoline (1 g.) was acetylated as in Section II (b). The acetyl derivative crystallized from ethyl acetate as small yellow prisms (0.94 g.), m.p. and mixed m.p. with xanthoxoline acetate 238–240 °C. The synthetic compound gave the same colour reactions as the acetyl derivative of the natural product.

Found: C, 65.0; H, 4.8; N, 4.4%.

Calculated for $\text{C}_{17}\text{H}_{15}\text{O}_5\text{N}$: C, 65.2; H, 4.8; N, 4.5%.

III. REFERENCES

- CAMPS, R. (1901).—*Ber. dtsh. chem. Ges.* **34**: 2714.
CROW, W. D., and PRICE, J. R. (1949).—*Aust. J. Sci. Res. A* **2**: 249.
HUGHES, G. K., MATHESON, N. K., NORMAN, A. T., and RITCHIE, E. (1952).—*Aust. J. Sci. Res. A* **5**: 206.
HUGHES, G. K., and NEILL, K. G. (1949).—*Aust. J. Sci. Res. A* **2**: 429.
HUGHES, G. K., NEILL, K. G., and RITCHIE, E. (1950).—*Aust. J. Sci. Res. A* **3**: 497.
HUGHES, G. K., NEILL, K. G., and RITCHIE, E. (1952).—*Aust. J. Sci. Res. A* **5**: 401.

THE CHEMICAL CONSTITUENTS OF AUSTRALIAN *FLINDERSIA* SPECIES

III. THE ALKALOIDS OF *FLINDERSIA* COLLINA BAIL.

By F. A. L. ANET,* P. T. GILHAM,* P. GOW,* G. K. HUGHES,*
and E. RITCHIE*

[Manuscript received November 26, 1951]

Summary

The known alkaloid kokusaginine and a new alkaloid *flindersiamine* have been isolated from *F. collina* Bail. They have been shown to be 6,7-dimethoxy- and 8-methoxy-6,7-methylenedioxy-derivatives respectively of dictamnine.

I. INTRODUCTION

In Part I of this series (Anet, Hughes, and Ritchie 1949) mention was made of the isolation of two alkaloids from the bark of *F. collina* Bail. and in this paper a study of their constitutions is presented. The first of these has been identified as kokusaginine, originally isolated by Terasaka (1933). The other, a new alkaloid, has been called *flindersiamine* (not to be confused with flindersine an alkaloid isolated by Mathes and Schreiber (1944) from *F. australis* R.Br.) as it has been found in other *Flindersia* species. These latter occurrences will be reported in subsequent communications.

II. THE STRUCTURE OF KOKUSAGININE

Previous work by Terasaka (loc. cit.) showed that kokusaginine had the composition $C_{13}H_{14}O_4N$ and contained three methoxyl groups. On heating with methyl iodide it isomerizes to the very weakly basic *isokokusaginine* which now contains only two methoxyl groups. The absorption spectrum was similar to that of certain furoquinoline alkaloids which also show the above behaviour. In the present study these results were confirmed.

Further, oxidation with potassium permanganate in acetone gave an aldehyde $C_{13}H_{13}O_5N$ (also obtained by ozonolysis) and the related acid $C_{13}H_{13}O_6N$ which on boiling with dilute hydrochloric acid lost a molecule of carbon dioxide and a methyl group and gave a high melting alkali-soluble substance $C_{11}H_{11}O_4N$ which contained two methoxyl groups. This gave a dark red nitroso-compound, a reaction characteristic of 2,4-dihydroxyquinolines, the expected type of product from a furoquinoline alkaloid.

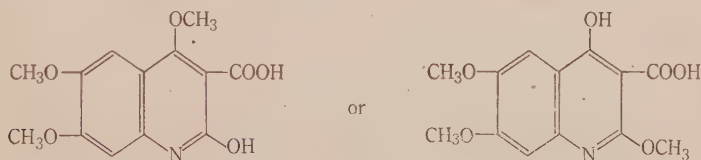
Assuming this, the position of the two methoxyl groups was proved as follows. As no demethylation of *isokokusaginine* took place with hydrochloric acid at 100°C. it follows that a 5-methoxy-group is absent (compare the nor-

* Department of Organic Chemistry, University of Sydney.

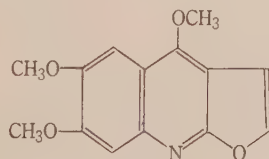
position in the acridone alkaloids (Crow and Price 1949) and the ready mono-demethylation of acronycidine (Lahey, Lamberton, and Price 1950)). However, *isokokusaginine* was demethylated by hydrobromic acid in acetic acid to a dihydric phenol (characterized as a *diacetate*) which gave an intense blue-violet colour with ferric chloride, indicating the presence of two adjacent hydroxyl groups. This suggested that the methoxyl groups in *isokokusaginine* were attached to the 6,7-positions since, the only alternative, the 7,8-derivative, is already known, being *isoskimmianine* (Asahina and Inabuse 1930).

Asahina and Nakanishi (1930) had already prepared 2,4-dihydroxy-6,7-dimethoxyquinoline and described it as melting above 300 °C. Repetition of their work gave a product, m.p. and mixed m.p. with the substance derived from kokusaginine 350–355 °C. The nitroso-compounds, which are difficult to purify, also appeared identical. As neither of these substances was considered satisfactory for comparison (acetates were also discarded) a methyl derivative was prepared from each of m.p. and mixed m.p. 216–218 °C. It had the composition $C_{13}H_{15}O_4N$ and was believed to be 4,6,7-trimethoxy-1-methylcarbostyryl. X-ray powder photographs supported the chemical evidence.

This still allows a linear or angular structure for kokusaginine. The former is favoured as dictamnine has been shown to have that structure. The following evidence supports this contention. The acid from the oxidation of kokusaginine may have either of the two structures



the former being derived from the linear and the latter from the angular structure. The latter, synthesized by the reductive cyclization of dimethyl 6-nitroveratroyl-malonate under anhydrous conditions followed by hydrolysis with potassium hydroxide, was not the same as the oxidation product. Hence kokusaginine is 6,7-dimethoxydictamnine (I)

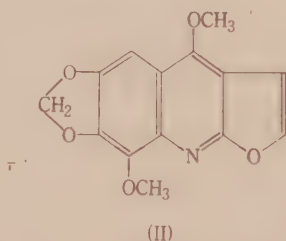


(I)

III. THE STRUCTURE OF FLINDERSIAMINE

Flindersiamine $C_{14}H_{11}O_5N$ contains two methoxyl groups and gives a positive methylenedioxy test. Its ultraviolet absorption spectrum resembles that of kokusaginine (I). On heating with methyl iodide it gives *isoflindersiamine* a weak base with only one methoxyl group.

Oxidation with potassium permanganate in acetone gave an *acid* $C_{13}H_{11}O_7N$. Boiling this acid with dilute hydrochloric acid afforded two products, the expected *methoxymethylenedioxy-2,4-dihydroxyquinoline* $C_{11}H_9O_5N$ and a second substance which will be discussed later. The dihydroxyquinoline was unaffected by boiling hydrochloric acid and is therefore unlikely to have a methoxy- or methylenedioxy-group attached to position 5. This allows only two alternative structures viz. 6-methoxy-7,8-methylenedioxy- and 8-methoxy-6,7-methylenedioxy-2,4-dihydroxyquinoline. The latter, which was more readily available, was synthesized and found to be identical with the degradation product. Comparison of the *nitroso*- and *acetyl* derivatives confirmed their identity. X-ray powder photographs of the acetates showed no differences. Thus flindersiamine is 8-methoxy-6,7-methylenedioxydictamnine (II)



Some improvements in the preparation of 6-nitromyristicine acid are noted in the experimental section.

A by-product obtained in the preparation of the above dihydroxyquinoline showed the same properties as the second degradation product of the acid obtained by the oxidation of flindersiamine. Neither gave a nitroso-compound, both could be sublimed unchanged, and both on boiling with dilute sodium hydroxide were converted into the dihydroxyquinoline. Analytical figures did not allow a formula to be proposed but work on this substance is proceeding.

IV. EXPERIMENTAL

Melting points are uncorrected.

(a) Isolation of the Alkaloids

For the isolation of the alkaloids the following method gave the best results. The finely milled bark (1 kg.) was exhausted with chloroform, which after concentration was thoroughly extracted with 5% hydrochloric acid. The combined acid extracts were basified, the precipitate transferred to chloroform which was filtered through a layer of kieselguhr, and evaporated to dryness. The residue (2 g.) was dissolved in hot 2% hydrochloric acid (100 cc.) decolourized with charcoal and cooled to 0°C. The crystalline solid (0.5 g.) which separated consisted mainly of kokusaginine hydrochloride. This after one recrystallization from dilute hydrochloric acid was converted to the base, which was recrystallized from a small volume of chloroform and then from methanol as colourless needles (0.3 g.), m.p. 168–169°C.

Found: C, 64.2; H, 5.2; N, 5.5; CH_3O , 36.3%.

Calculated for $C_{14}H_{13}O_4N$: C, 64.2; H, 5.1; N, 5.4; $3 \times CH_3O$, 35.9%.

The aqueous filtrates from above were basified and extracted with chloroform. The extract on concentration and cooling yielded crude flindersiamine which after repeated crystallization

from methanol was obtained as thick white needles (1.0 g.), m.p. 206–207 °C. From benzene it separated as small cubes.

Found: C, 61.2; H, 4.3; N, 5.2; CH_3O , 22.3%.

Calculated for $\text{C}_{14}\text{H}_{11}\text{O}_5\text{N}$: C, 61.5; H, 4.1; N, 5.2; $2 \times \text{CH}_3\text{O}$, 22.7%.

Neither of the alkaloids fluoresced in ultraviolet light.

(b) *Kokusaginine Picrate*

This was prepared in alcoholic solution and recrystallized from alcohol as yellow needles, m.p. 218–219 °C. (lit. 218.5–219.5 °C.).

(c) *isoKokusaginine*

This was prepared as described by Terasaka (loc. cit.). It became pink on exposure to light and gradually decomposed on long standing. It was recrystallized from methanol, m.p. 249–250 °C. (lit. 248–249 °C.).

Found: C, 64.1; H, 5.1; N, 5.4; CH_3O , 24.2%.

Calculated for $\text{C}_{14}\text{H}_{13}\text{O}_4\text{N}$: C, 64.2; H, 5.1; N, 5.4; $2 \times \text{CH}_3\text{O}$, 24.0%.

(d) *Permanganate Oxidation of Kokusaginine*

To kokusaginine (3 g.) in acetone (400 cc.) finely powdered potassium permanganate (4.5 g.) was gradually added with vigorous stirring. The brown sludge was collected, thoroughly washed with dilute sodium hydroxide, and the aqueous extracts acidified. The gelatinous precipitate was collected, washed well with water, and recrystallized from ethanol (charcoal) as an off-white powder (0.8 g.). When gradually heated from room temperature it sintered at 295 °C. but did not melt below 350 °C. However, if placed in a bath at 340 °C., it melted, frothed slightly, and immediately solidified to a mass of pale yellow needles which melted again to a deep yellow liquid which rapidly darkened. It did not give any colour with ferric chloride.

Found: C, 55.6; H, 4.7; N, 5.1; CH_3O , 31.8%.

Calculated for $\text{C}_{13}\text{H}_{13}\text{O}_6\text{N}$: C, 55.9; H, 4.7; N, 5.0; $3 \times \text{CH}_3\text{O}$, 33.3%.

The aldehyde (0.3 g.) obtained from the acetone extracts was recrystallized from ethyl acetate as fine yellow needles, m.p. 248–249 °C. The melting point was not depressed by the product described in Section IV (e).

(e) *Ozonolysis of Kokusaginine*

Kokusaginine (1.0 g.) in anhydrous ethyl acetate (300 cc.) was cooled in an ice salt bath and ozone (6%) passed in for 15 minutes. Palladized calcium carbonate (0.5 g.) was added and the mixture shaken with hydrogen at atmospheric pressure for 3 hours. After removal of the catalyst, the solution was concentrated and the precipitate recrystallized from ethyl acetate as fine pale yellow needles (0.38 g.), m.p. 251–252 °C.

Found: C, 59.5; H, 4.9%.

Calculated for $\text{C}_{13}\text{H}_{13}\text{O}_5\text{N}$: C, 59.3; H, 5.0%.

(f) *Degradation of the Acid $\text{C}_{13}\text{H}_{13}\text{O}_6\text{N}$*

The acid (0.28 g.) was heated under reflux with 15% hydrochloric acid (40 cc.). The clear solution gradually deposited fine white needles which were collected after 2 hours. After recrystallization from methanol or glacial acetic acid a product (0.16 g.) was obtained, m.p. 350–355 °C. This compound sublimed at 250 °C./ 10^{-3} mm. and readily absorbed moisture. For analysis it was dried at 117 °C. *in vacuo*.

Found: C, 59.3; H, 5.4; N, 6.4; CH_3O , 27.8%.

Calculated for $\text{C}_{11}\text{H}_{11}\text{O}_4\text{N}$: C, 59.7; H, 5.0; N, 6.3; $2 \times \text{CH}_3\text{O}$, 28.1%.

It gave a dark red nitroso-compound, which was difficult to purify, m.p. 264–266 °C. (decomp.).

(g) *Demethylation of isoKokusaginine*

isoKokusaginine (0.5 g.), glacial acetic acid (5 cc.), and hydrobromic acid (10 cc., d. 1.7) were heated under reflux for 4 hours. The solution was cooled to 0 °C. and the precipitated hydrobromide collected. As the free base was difficult to purify it was converted to the acetate. This was recrystallized from aqueous methanol as white prisms, m.p. 240–242 °C.

Found: C, 60.6; H, 4.5%.

Calculated for $C_{16}H_{13}O_3N$: C, 61.0; H, 4.2%.

The phenol gave an intense blue-violet colour with ferric chloride in both aqueous and alcoholic solutions.

(h) *2,4-Dihydroxy-6,7-dimethoxyquinoline*

This was prepared by the method of Asahina and Nakanishi (loc. cit.) except that the intermediate nitrobenzoyl malonic ester was obtained as follows.

Ethoxymagnesium malonic ester, from magnesium turnings (0.2 g.), dry alcohol (1.2 cc.), and diethyl malonate (1.2 cc.), in ethereal solution was added slowly to a suspension of 6-nitroveratroyl chloride (from 2.0 g. of acid) in ether. After the vigorous reaction had subsided the mixture was refluxed for 1 hour, cooled, and made acid with dilute sulphuric acid. The ethereal layer was evaporated and the residual red oil refluxed with tin and hydrochloric acid.

The product (0.7 g.) after repeated recrystallization from alcohol had m.p. 350–355 °C. undepressed on admixture with the product derived from kokusaginine. The nitroso-compound was a dark red powder, m.p. 264–266 °C. (decomp.) (lit. 266 °C. (decomp.)).

(i) *4,6,7-Trimethoxy-1-methylcarbostyrl*

The above dihydroxyquinoline (0.5 g.) was dissolved in an excess of 25% sodium hydroxide and treated gradually, at room temperature, with a large excess of dimethyl sulphate (5 g.). The alkaline solution was extracted with chloroform and, after drying, the chloroform solution was passed through a column of alumina. The eluate was evaporated and the residue recrystallized from benzene-light petroleum as colourless cubes (0.2 g.), m.p. 216–218 °C.

Found: C, 62.7; H, 6.0; N, 5.7%.

Calculated for $C_{13}H_{16}O_4N$: C, 62.7; H, 6.0; N, 5.6%.

The above method was first applied to 2,4-dihydroxyquinoline and gave fine white needles, m.p. 68–69 °C. Friedlander and Muller (1887) record 68–69 °C. for 4-methoxy-1-methylcarbostyrl made by another method.

By the same method using the product from kokusaginine a substance, m.p. 216–218 °C., undepressed by the above, was obtained.

Found: N, 5.6%.

Calculated for $C_{13}H_{16}O_4N$: N, 5.6%.

(j) *4-Hydroxy-2,6,7-trimethoxyquinoline-3-carboxylic Acid*

Dimethyl 6-nitroveratroylmalonate, prepared by the method of Asahina and Nakanishi (loc. cit.) as a viscous red oil (3 g.), granulated zinc (10 g.), and dry methanol (30 cc.) was saturated with dry hydrogen chloride until all the zinc had dissolved. The solution was neutralized with 10% sodium bicarbonate and then made just acid with dilute acetic acid. The amorphous brown precipitate was collected and hydrolysed by standing overnight in 5% methanolic potassium hydroxide (20 cc.).

The acid obtained on acidification was recrystallized from glacial acetic acid as small colourless needles of indefinite m.p. If placed in a bath at 330 °C. it melted with vigorous frothing and resolidified to a black mass. Its solution in alcohol gave a deep reddish brown colour with ferric chloride.

Found: C, 55.7; H, 4.7; N, 5.1; CH_3O , 32.7%.

Calculated for $C_{13}H_{13}O_6N$: C, 55.9; H, 4.7; N, 5.0; $3 \times CH_3O$, 33.3%.

(k) *Flindersiamine*

On warming flindersiamine with concentrated sulphuric acid it gave first a transient pink and then a deep blue colour which finally changed to green. This test is very sensitive and is also given by *isoflindersiamine* but not by any of the degradation products where the furane ring is destroyed. In these latter substances a positive methylenedioxy test was obtained either with gallic acid or kokusaginine, and hence it would appear that the above test with flindersiamine and sulphuric acid is in fact a methylenedioxy test similar to the Labat test. *Flindersiamine picrate* was prepared in propyl alcohol and recrystallized from the same solvent as yellow needles, m.p. 200–201 °C.

Found: N, 11.1%.

Calculated for $C_{17}H_{14}O_{11}N_4$: N, 11.2%.

(l) *isoflindersiamine*

Flindersiamine was heated with a large excess of methyl iodide for 4 hours at 100 °C. The product was crystallized from aqueous methanol, m.p. 209–211 °C.

Found: C, 53.8; H, 4.9; N, 4.6; CH_3O , 9.6%.

Calculated for $C_{14}H_{11}O_5N \cdot 2H_2O$: C, 54.4; H, 4.9; N, 4.5; $1 \times CH_3O$, 10.0%.

(m) *Oxidation of Flindersiamine*

This was carried out as for kokusaginine using flindersiamine (5 g.), acetone (500 cc.), and potassium permanganate (9 g.). After several recrystallizations from glacial acetic acid the product (1.0 g.) was obtained as colourless needles, m.p. 249–250 °C. It did not give a colour with ferric chloride.

Found: C, 52.9; H, 4.1; N, 5.1%.

Calculated for $C_{13}H_{11}O_7N$: C, 53.2; H, 3.8; N, 4.8%.

(n) *Acid Degradation*

The above acid (1 g.) was refluxed with 10% hydrochloric acid (500 cc.) for 4 hours, cooled, and the precipitate collected and recrystallized from glacial acetic acid as colourless needles, m.p. 295 °C. (decomp.). It sublimed at 250 °C./ 10^{-3} mm.

This substance did not give a nitroso-derivative but after boiling with sodium hydroxide for 10 minutes a deep red nitroso-compound was obtained on addition of sodium nitrite and hydrochloric acid. The alkaline solution yielded the expected dihydroxyquinoline on acidification.

The original acid filtrate was evaporated to dryness, the residue treated with a little water, and the product collected and recrystallized from glacial acetic acid as stout needles, m.p. 290–295 °C. after sintering at 280 °C.

Found: C, 56.2; H, 4.0; N, 5.9; CH_3O , 12.9%.

Calculated for $C_{11}H_9O_5N$: C, 56.2; H, 3.9; N, 6.0; $1 \times CH_3O$, 13.2%.

The nitroso-derivative prepared as above after recrystallization from glacial acetic formed red rosettes of needles of indefinite m.p. 240–245 °C.

Found: N, 10.6%.

Calculated for $C_{11}H_8O_6N_2$: N, 10.6%.

(o) *Demethylation of isoflindersiamine*

As for *isokokusaginine* the product (yield 60%) recrystallized from ethanol, m.p. 275–280 °C. (decomp.).

Found: C, 60.3; H, 4.1; N, 5.8; CH_3O , 0%.

Calculated for $C_{13}H_9NO_5$: C, 60.3; H, 3.8; N, 5.5%.

A suspension of the phenol in ether treated with diazomethane gave *isoflindersiamine*, m.p. and mixed m.p. 208–209 °C.

(p) *2,4-Dihydroxy-8-methoxy-6,7-methylenedioxyquinoline*

The myristicin aldehyde used was made by the methylenation of gallaldehyde-3-methyl ether (Bradley, Robinson, and Swarzenbach 1930) using the method of Rao and Seshadri (1946). It was found that vigorous stirring at a temperature not exceeding 60 °C. was essential for a satisfactory yield. Oxidation with permanganate gave the acid which was converted to the methyl ester with diazomethane in almost quantitative yield.

The methyl ester was nitrated as described by Salway (1911) for the ethyl ester and the product was hydrolysed to the acid as follows.

The nitro-ester (5.6 g.), alcohol (25 cc.), and 4% sodium hydroxide (35 cc.) were warmed until solution was complete and then allowed to stand overnight. The pale yellow solution was made just acid and the cream precipitate recrystallized from alcohol as pale yellow needles (4.6 g.), m.p. 245 °C.

If the hydrolysis was carried out under reflux a deep red solution was obtained and the acid was difficult to purify.

6-Nitromyristicin acid (2 g.), chloroform (10 cc.), and redistilled thionyl chloride (3 cc.) were heated under reflux for 30 minutes. The solvent was removed *in vacuo* and the residue recrystallized from light petroleum giving the acid chloride as almost colourless needles (2 g.), m.p. 127–128 °C.

To ethoxymagnesium malonic ester, from magnesium turnings (0.26 g.), dry alcohol (0.25 cc.), and dry ether (3.7 cc.), was added with shaking an ethereal solution of the above acid chloride (2 g.). After refluxing for 30 minutes the mixture was cooled, made just acid with dilute sulphuric acid, and the ether layer separated. The aqueous layer was extracted twice with ether and the combined ether extracts evaporated to dryness. The resultant red gum was dissolved in alcohol (40 cc.), concentrated hydrochloric acid (40 cc.) and tin (20 g.) added, and the whole refluxed for 6 hours. On addition of water (200 cc.) and alcohol (200 cc.) a pale yellow compound separated. This was collected and recrystallized from glacial acetic acid as colourless needles (0.4 g.), m.p. 295 °C. (decomp.). It sublimed unchanged at 250 °C. 10⁻³ mm.

Found: C, 51.5; H, 3.1; N, 5.4%.

Calculated for C₁₃H₉O₇N: C, 51.6; H, 3.2; N, 5.0%.

It was sparingly soluble in most organic solvents and did not give a nitroso-derivative. However, on boiling with sodium hydroxide it was converted to the substance described below.

The aqueous alcoholic acid filtrate was saturated with hydrogen sulphide, the tin sulphides removed, and the light coloured filtrate evaporated to a small bulk, when white needles of a hydrochloride separated. These were collected, treated with water which gave the free base, and then crystallized from glacial acetic acid as colourless needles (0.1 g.), m.p. 295 °C. (after softening at 280 °C.) undepressed on admixture with the substance derived from flindersiamine.

Found: C, 56.3; H, 4.2; N, 5.9; CH₃O, 12.9%.

Calculated for C₁₁H₉O₅N: C, 56.2; H, 3.9; N, 6.0; 1 × CH₃O, 13.2%.

An acetate, presumably the 4-acetyl derivative, prepared from each using acetic anhydride and pyridine was obtained as colourless needles, m.p. 231 °C., either separately or on admixture.

Found for the synthetic product: C, 56.7; H, 4.1; N, 5.3%.

Found for the degradation product: N, 5.2%.

Calculated for C₁₄H₁₁O₆N: C, 56.3; H, 4.0; N, 5.1%.

V. ACKNOWLEDGMENTS

The authors are indebted to Mr. L. J. Webb, C.S.I.R.O., for collection of the bark; to Mrs. E. Bielski, Miss J. Fildes, and Miss B. Naylor, for all the analyses; to the Dunlop Rubber Company for the award of a scholarship to one of them (F.A.L.A.); to Mr. R. W. Hinde who carried out several experiments on the ozonolysis of kokusaginine; and to Dr. D. P. Mellor for the X-ray powder photographs.

VI. REFERENCES

- ANET, F. A. L., HUGHES, G. K., and RITCHIE, E. (1949).—*Aust. J. Sci. Res. A* **2**: 127.
- ASAHINA, Y., and INABUSE, M. (1930).—*Ber. dtsh. chem. Ges.* **63**: 2052.
- ASAHINA, Y., and NAKANISHI, S. (1930).—*Ber. dtsh. chem. Ges.* **63**: 2057.
- BRADLEY, W., ROBINSON, R., and SWARZENBACH, G. (1930).—*J. Chem. Soc.* **1930**: 811.
- CROW, W. D., and PRICE, J. R. (1949).—*Aust. J. Sci. Res. A* **2**: 282.
- FRIEDLANDER, P., and MÜLLER, F. (1887).—*Ber. dtsh. chem. Ges.* **20**: 2009.
- LAHEY, F. N., LAMBERTON, J. A., and PRICE, J. R. (1950).—*Aust. J. Sci. Res. A* **3**: 155.
- MATHES, H., and SCHREIBER, E. (1914).—*Ber. dtsh. Pharm. Ges.* **24**: 385.
- RAO, K. V., and SESHADRI, T. R. (1946).—*Proc. Indian Acad. Sci.* **23**: 147.
- SALWAY, A. (1911).—*J. Chem. Soc.* **99**: 266.
- TERASAKA, M. (1933).—*J. Pharm. Soc. Japan* **53**: 219. (*Chem. Abstr.* **21**: 7336 (1935).)

SHORT COMMUNICATIONS

THE CHEMICAL CONSTITUENTS OF AUSTRALIAN *FLINDERSIA* SPECIES

IV. THE CONSTITUENTS OF *FLINDERSIA* *BOURJOTIANA* F. MUELL.*

By J. R. CANNON,† G. K. HUGHES,† J. R. PRICE,‡ and E. RITCHIE†

Flindersia bourjotiana F. Muell. is a large tree found in the rain-forest of the Atherton Tableland, Queensland. It attains a height of about 120 ft. and a stem diameter of up to 30 in. The standard timber name is Queensland silver ash. When the bark is cut it is pinkish brown in colour becoming yellower towards the sapwood.

From a concentrated methanolic extract of the bark, a high yield of a colourless, crystalline substance, m.p. 256–257° C. (decomp.) was obtained and identified as the well-known glycoside hesperidin from its constants and by comparison with an authentic specimen. The extract also yielded a small amount of colourless, crystalline, basic product, m.p. 182–183° C. This melting point was unaltered on repeated crystallization from methanol, and a paper partition chromatogram gave only a single spot, but further examination revealed that the product was a mixture of skimmianine and flindersiamine (Anet *et al.* 1952).

Experimental

All melting points are uncorrected. Analyses are by Dr. Weiler and Dr. Strauss, Oxford.

(i) *Extraction of the Bark and Isolation of Hesperidin*.—The milled bark was boiled with methanol repeatedly, the extracts filtered hot and evaporated until crystallization commenced. The crude glycoside (yield 1.4%) was filtered off and washed with a large volume of methanol. The substance was best purified by dissolving it in a large volume of boiling methanol, then evaporating the solution until crystallization began, when it was obtained as colourless micro-needles, m.p. and mixed m.p. with an authentic specimen of hesperidin 256–257° C. (decomp.), $[\alpha]_D^{30} -77.5^\circ$ (c 0.428 in pyridine) (lit. $[\alpha]_D^{20} -75.8^\circ$).

The acetate crystallized from methanol as colourless micro-needles, m.p. and mixed m.p. with hesperidin octa-acetate 175–176° C., $[\alpha]_D^{21} -45.8^\circ$ (c 1.038 in pyridine) (lit. $[\alpha]_D^{20} -47.3^\circ$).

The aglycone, hesperitin crystallized from aqueous alcohol as needles, m.p. 227–229° C. (decomp.) (lit. 226° C.); its acetate crystallized from alcohol as colourless needles, m.p. 126–127° C. (lit. 127–129° C.).

(ii) *Isolation of the Alkaloids*.—The methanolic filtrate from the hesperidin was evaporated on the water-bath and the residue exhaustively extracted with chloroform. The chloroform solution was shaken repeatedly with 10% hydrochloric acid, the acid extracts basified with ammonia, and the crude product dissolved in benzene. This solution was run through a short

* Manuscript received November 26, 1951.

† Department of Organic Chemistry, University of Sydney.

‡ Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

column of alumina to remove coloured impurities, the pale yellow eluate evaporated, and the residue crystallized from methanol as colourless prisms (yield 0.03%), m.p. 182–183 °C. This product gave a single spot (with iodine) on a paper partition chromatogram using *n*-butanol-5% acetic acid, and the melting point was not raised on recrystallization from methanol or benzene. On warming with concentrated sulphuric acid the product gave a prussian blue coloured solution.

A partial separation of the component alkaloids was effected by extracting a solution of the product (2 g.) in benzene (200 ml.) with 10 ml. portions of 0.5% hydrochloric acid until practically no precipitate (*A*) was obtained on basifying the acid solution. The remainder of the alkaloids (*B*) in the benzene layer was extracted with 5% hydrochloric acid. Each of these acid extracts was treated with ammonia and the precipitated bases dissolved in the minimum volume of benzene. The solutions were then run through columns of alumina which were sufficiently long to adsorb completely all basic material, and each column was developed with benzene. Every 25 ml. of eluate was evaporated and the m.p. of each residue and the colour reaction with hot concentrated sulphuric acid was determined.

It was found that the first fractions from *A* gave a strongly positive colour test (prussian blue) and melted 180–185 °C. As the chromatogram was developed the colour test grew much weaker and the m.p. of the residue fell to 176–178 °C. These last fractions were combined and recrystallized from methanol when colourless prisms were obtained, m.p. and mixed m.p. with an authentic specimen of skimmianine from *Melicope fareana* F. Muell. (Crow and Price 1949) 175–176 °C. A concentrated sulphuric acid solution gave only a very pale green colour on warming, identical with the colour obtained with authentic skimmianine.

Found: N, 5.2; CH_3O , 35.6%.

Calculated for $\text{C}_{14}\text{H}_{13}\text{O}_4\text{N}$: N, 5.4; $3 \times \text{CH}_3\text{O}$, 35.9%.

The picrate formed flat, yellow needles from methanol, m.p. and mixed m.p. 192–194 °C.

The first benzene eluates from the column containing *B* had m.p. 204–208 °C. and gave a very strong colour test. The m.p. of the later fractions gradually fell to about 190 °C. Those fractions of m.p. above 200 °C. were combined and crystallized from alcohol when colourless needles were obtained, m.p. and mixed m.p. with authentic flindersiamine from *Flindersia collina* 207–208 °C. The colourless solution in concentrated sulphuric acid became intensely prussian blue on warming.

Found: C, 61.4; H, 3.9%.

Calculated for $\text{C}_{14}\text{H}_{11}\text{O}_5\text{N}$: C, 61.5; H, 4.1%.

The picrate crystallized as bright yellow laths from *n*-propanol, m.p. and mixed m.p. 200 °C. (decomp.).

Found: N, 11.1%.

Calculated for $\text{C}_{20}\text{H}_{14}\text{O}_{12}\text{N}_4$: N, 11.2%.

(iii) *Separation of isoSkimmianine and isoFlindersiamine.*—The product, m.p. 182–183 °C. (2 g.) in chloroform (1 ml.) was heated with methyl iodide (3 ml.) in a sealed tube at 100 °C. for 4 hours. The solution was evaporated to dryness and the residue extracted with cold benzene. The solution (*C*) was filtered from the insoluble residue (*D*) and chromatographed on alumina. Evaporation of the eluate and recrystallization of the residue from water yielded colourless needles, m.p. and mixed m.p. with *isoskimmianine* 184–185 °C. *isoSkimmianine* also gave a pale green solution on heating with concentrated sulphuric acid.

Found: C, 64.7; H, 5.3; N, 5.5%.

Calculated for $\text{C}_{14}\text{H}_{13}\text{O}_4\text{N}$: C, 64.9; H, 5.1; N, 5.4%.

Crystallization of *D* from benzene, then from aqueous alcohol, yielded fine colourless needles, m.p. and mixed m.p. with *isoflindersiamine* 208–209 °C. (decomp.); mixed m.p. with flindersiamine 160–170 °C. *isoFlindersiamine* gave a prussian blue coloured solution on warming with concentrated sulphuric acid.

Found: N, 5.0%.

Calculated for $\text{C}_{14}\text{H}_{11}\text{O}_5\text{N}$: N, 5.1%.

(iv) *Examination of a Synthetic Mixture of Skimmianine and Flindersiamine.* Approximately equal quantities of authentic skimmianine and flindersiamine were dissolved in hot methanol, the solution seeded with a crystal of the naturally-occurring mixture, and allowed to cool. Large, characteristic prisms were obtained, m.p. 183–184° C., which did not depress the m.p. of the natural product.

The authors wish to thank Mr. L. J. Webb, C.S.I.R.O., who collected and identified the bark used in this investigation.

References

- ANET, F. A. L., GILHAM, P. T., GOW, P., HUGHES, G. K., and RITCHIE, E. (1952).—*Aust. J. Sci. Res. A* **5**: 412.
CROW, W. D., and PRICE, J. R. (1949).—*Aust. J. Sci. Res. A* **2**: 249.

A SIMPLIFIED FORM OF THE RELATIVISTIC ELECTROMAGNETIC EQUATIONS

By N. W. TAYLOR*

[Manuscript received March 24, 1952]

Summary

It is shown how Maxwell's electromagnetic equations in General Relativity may be expressed in the form of a single four-vector density equation, in which the field tensor has only three distinct complex components. The number of equations is reduced, but all the usual classical formulae may be obtained by separating the real and imaginary parts.

I. MAXWELL'S EQUATIONS AND THE FIELD COMPONENTS

Maxwell's equations of the electromagnetic field, in their most fundamental form ($\mathbf{D}=\mathbf{E}$, $\mathbf{B}=\mathbf{H}$), may be represented by the single expression

$$\frac{\partial}{\partial x^\nu}(\mathbf{F}^{\mu\nu})=\mathbf{J}^\mu\dagger, \quad \dots\dots\dots (1)$$

where \mathbf{J}^μ is the usual relativistic current density

$$\mathbf{J}^\mu = -\frac{1}{c}(j_1, j_2, j_3, ic), \quad \dots\dots\dots (2)$$

and the field tensor is given by

$$\mathbf{F}^{\mu\nu} = \frac{1}{\sqrt{g}}(\mathbf{F}^{\mu\nu}), \quad \dots\dots\dots (3)$$

which has only three distinct components instead of the usual six, reducing in geodesic systems of coordinates which are locally Cartesian ($g_{\mu\nu}=\delta_{\mu\nu}$ locally) to the following form :

$$\left. \begin{array}{cccc} \mathbf{F}^{\mu\nu} = & 0 & F_3 & -F_2 & F_1 \\ & -F_3 & 0 & F_1 & F_2 \\ & F_2 & F_1 & 0 & F_3 \\ & -F_1 & -F_2 & -F_3 & 0 \end{array} \right\}, \quad \dots\dots (4)$$

F_1 being given in terms of the classical field components E_1 and H_1 by

$$F_1 = iE_1 - H_1, \quad \dots\dots\dots (5)$$

with similar expressions for F_2 and F_3 .

* Department of Mathematics, New England University College, Armidale, N.S.W.

† In this paper roman type shall denote tensor densities, and italic type shall refer to ordinary tensors.

That the equations (1) do in fact give Maxwell's equations may be shown in the usual manner, by writing out the components in the Cartesian framework (x, y, z, ict), the two sets of equations being given, in this case, by equating real and imaginary parts. From the real part we have

$$\frac{\partial \mathbf{E}}{\partial t} + \mathbf{j} = c \operatorname{curl} \mathbf{H},$$

$$\operatorname{div} \mathbf{H} = 0,$$

and from the imaginary part,

$$\frac{\partial \mathbf{H}}{\partial t} = -c \operatorname{curl} \mathbf{E},$$

$$\operatorname{div} \mathbf{E} = \rho.$$

The equations will not be relativistic unless we can find an invariant condition governing the form of the field tensor such that it reduces to a system of the form (4) in all geodesic systems of coordinates in which $g^{\mu\nu} = \delta^{\mu\nu}$, locally. In the usual theory the field tensor is restricted by the condition of antisymmetry, having six different components. The necessary and sufficient condition for this to apply in all systems of coordinates is expressible in the form of the tensor equation

$$F^{\mu\nu} = -F^{\nu\mu}.$$

In the present theory, however, there is a greater restriction on the components of the field tensor, and the above condition, although necessary, is no longer sufficient. Consider the tensor-density equation

$$F^{\mu\nu} = \frac{1}{2} \varepsilon^{\mu\nu\sigma\tau} F_{\sigma\tau}, \quad \dots\dots\dots (4')$$

where $\varepsilon^{\mu\nu\sigma\tau}$ is the tensor-density

$$\varepsilon^{\mu\nu\sigma\tau} = +1 \text{ if } \mu\nu\sigma\tau \text{ is an even permutation of } 1234,$$

$$-1 \text{ if it is an odd permutation,}$$

$$0 \text{ in all other cases.}$$

In the particular geodesic coordinates this reduces to

$$F^{\mu\nu} = \frac{1}{2} \varepsilon^{\mu\nu\sigma\tau} F_{\sigma\tau},$$

leading to (4) for all such systems. For example,

$$F^{12} = \frac{1}{2} (F^{34} - F^{43}) = F^{34} = F_3.$$

Therefore, by the Principle of Covariance and the Principle of Equivalence, (4) must be replaced by (4') if general coordinate systems are under investigation. The identification of the field tensor with the classical field components must, as usual, be made in a special framework, so that there is no covariant equation corresponding to (5).

The Lorentz-Einstein equations for the transformation of the electric and magnetic field components are obtained as before by considering the equations

$$F'^{\mu\nu} = \frac{\partial x'^{\mu}}{\partial x^{\sigma}} \frac{\partial x'^{\nu}}{\partial x^{\tau}} F^{\sigma\tau},$$

making transformations only between the frameworks of special relativity. In this case, however, we need consider only three, instead of six equations

since we deal with an \mathbf{E} and \mathbf{H} component at the same time, and then separate them by equating real and imaginary parts.

II. THE POTENTIAL FUNCTION

In the original theory the six field components are not completely independent, but can be expressed in terms of a four-vector potential function, κ_μ , a second order tensor satisfying the condition of antisymmetry being constructed from the components of the derived function $\partial\kappa_\mu/\partial x^\nu$. Similarly, in the present theory, a potential function exists, but the components of the derivative must now be employed to construct a function which satisfies (4'). Such a function is

$$F^{\mu\nu} = (\sqrt{g} g^{\mu\sigma} g^{\nu\rho} + \frac{1}{2} \varepsilon^{\mu\nu\sigma\rho}) \left(\frac{\partial\kappa_\sigma}{\partial x^\rho} - \frac{\partial\kappa_\rho}{\partial x^\sigma} \right). \quad (6)$$

This equation satisfies the Principle of Covariance, and if the components are written out in the geodesic coordinates previously used, it gives a system with three different components, corresponding term by term with the array shown in (4).

We shall now investigate equation (6) in the framework of Special Relativity, supposing that

$$\kappa_\mu = (A_1, A_2, A_3, iV), \quad (7)$$

where $\mathbf{A} = (A_1, A_2, A_3)$ is the classical vector potential, and V the scalar potential. We have, for example,

$$F^{14} = F^{23} = \frac{\partial\kappa_1}{\partial x^4} - \frac{\partial\kappa_4}{\partial x^1} + \frac{\partial\kappa_2}{\partial x^3} - \frac{\partial\kappa_3}{\partial x^2}$$

from (6). Hence, from (4), (5), and (7)

$$\begin{aligned} F_1 &= iE_1 - H_1 = \frac{\partial A_1}{ic\partial t} - \frac{i\partial V}{\partial x} + \frac{\partial A_2}{\partial z} - \frac{\partial A_3}{\partial y}, \\ &= \{-i(\dot{\mathbf{A}}/c + \text{grad } V) - \text{curl } \mathbf{A}\}_x. \end{aligned}$$

Treating the other components similarly, and equating real and imaginary parts, we have the two classical equations:

$$\begin{aligned} \mathbf{H} &= \text{curl } \mathbf{A}, \\ \mathbf{E} &= -\text{grad } V - \dot{\mathbf{A}}/c. \end{aligned}$$

Hence the assumption (7) is correct.

Equation (6) does not determine κ_μ uniquely, since it is obvious that if κ_μ is a solution, so also is $\kappa_\mu + \partial\psi/\partial x^\mu$. This arbitrariness is restricted by taking

$$\partial\kappa_\mu/\partial x^\mu = 0, \quad (8)$$

which gives another classical equation

$$\text{div } \mathbf{A} + \dot{V}/c = 0.$$

Substituting from (6) in (1) we obtain the wave-equation

$$\frac{\partial}{\partial x} \left\{ \sqrt{g} g^{\mu\sigma} g^{\nu\rho} \left(\frac{\partial\kappa_\sigma}{\partial x^\rho} - \frac{\partial\kappa_\rho}{\partial x^\sigma} \right) \right\} = J^\mu. \quad (9)$$

In Cartesian systems this becomes, since $g_{\mu\nu} = \delta_{\mu\nu}$,

$$\frac{\partial}{\partial x^\nu} \left(\frac{\partial \kappa_\mu}{\partial x^\nu} - \frac{\partial \kappa_\nu}{\partial x^\mu} \right) = J^\mu,$$

or, by (8)

$$\frac{\partial^2 \kappa_\mu}{\partial x^\nu \partial x^\nu} = J^\mu.$$

In free space this gives

$$\frac{\partial^2 \kappa_\mu}{\partial x^\nu \partial x^\nu} = \frac{1}{c^2} \frac{\partial^2 \kappa_\mu}{\partial t^2}.$$

Sometimes, instead of the potential κ_μ , the corresponding third order antisymmetrical tensor-density

$$A^{\mu\nu\sigma} = \varepsilon^{\mu\nu\sigma\tau} \kappa_\tau, \quad (10)$$

is employed. In place of (6) we therefore have

$$F^{\mu\nu} = \frac{\partial A^{\mu\nu\sigma}}{\partial x^\sigma} + \frac{1}{2\sqrt{g}} \varepsilon^{\mu\nu\sigma\rho} g_{\sigma\alpha} g_{\rho\beta} \frac{\partial A^{\alpha\beta\tau}}{\partial x^\tau}, \quad (11)$$

a result which can be deduced from (6) and (10) by referring it to a geodesic system of coordinates.

III. THE CONSERVATION OF CHARGE

Since $F^{\mu\nu}$ is antisymmetric, we have from (1)

$$\frac{\partial J^\mu}{\partial x^\mu} = \frac{\partial^2 F^{\mu\nu}}{\partial x^\mu \partial x^\nu} = 0. \quad (12)$$

Expressing the current as a flow of charge, it can be written, in the framework of Special Relativity,

$$J^\mu = -\frac{1}{c} (\rho r_1, \rho r_2, \rho r_3, ic\rho). \quad (13)$$

This is a possible system of components since it is proportional to the velocity four-vector. It also denotes the transformation properties of the charge density ρ , leading to the result that the total charge on a given particle is a constant scalar. Substituting from (13) in (12) we have the classical equation of conservation:

$$\frac{\partial(\rho r_1)}{\partial x} + \frac{\partial(\rho r_2)}{\partial y} + \frac{\partial(\rho r_3)}{\partial z} + \frac{\partial\rho}{\partial t} = 0.$$

IV. THE PONDEROMOTIVE FORCE

Consider the contravariant vector

$$W^\mu = \frac{1}{g} F^{\mu\nu} J_\nu, \quad (14)$$

In the previous set of geodesic coordinates it may be written

$$W^\mu = F^{\mu\nu} A_\nu.$$

The x^1 component is

$$\begin{aligned} W^1 &= F^{12}J^2 + F^{13}J^3 + F^{14}J^4, \\ &= -\frac{1}{c}(F_3j_2 - F_2j_3) - F_1i\rho. \end{aligned}$$

Using (5), the real part of this is

$$R(W^1) = \frac{1}{c}(H_3j_2 - H_2j_3) + E_1\rho,$$

with similar expressions for the x^2 and x^3 components. Hence

$$R(W^1, W^2, W^3) = \frac{1}{c}\mathbf{j} \times \mathbf{H} + \rho\mathbf{E}.$$

In the same way we find for the imaginary part

$$I(W^1, W^2, W^3) = -\frac{1}{c}\mathbf{j} \times \mathbf{E} + \rho\mathbf{H}.$$

Consider now the x^4 component. It is

$$\begin{aligned} W^4 &= F^{41}J^1 + F^{42}J^2 + F^{43}J^3 \\ &= \frac{1}{c}(F_1j_1 + F_2j_2 + F_3j_3). \end{aligned}$$

The real and imaginary parts of this are

$$R(W^4) = -\frac{1}{c}\mathbf{j} \cdot \mathbf{H},$$

$$I(W^4) = \frac{1}{c}\mathbf{j} \cdot \mathbf{E}.$$

In constructing a four-vector from physical quantities, the space components are real and the time component imaginary, or vice versa. Hence we may write the complex W^μ in the form

$$W^\mu = P^\mu - iQ^\mu, \dots\dots\dots (15)$$

where both P^μ and Q^μ are four-vectors, such that in the coordinate frames under consideration

$$P^\mu = \left(\rho\mathbf{E} + \frac{1}{c}\mathbf{j} \times \mathbf{H}, \quad \frac{i}{c}\mathbf{j} \cdot \mathbf{E} \right),$$

$$Q^\mu = \left(\rho\mathbf{H} - \frac{1}{c}\mathbf{j} \times \mathbf{E}, \quad \frac{i}{c}\mathbf{j} \cdot \mathbf{H} \right).$$

The components of P^μ are therefore the force per unit volume and the power per unit volume. At the same time we have discovered another four-vector which is not sought in the usual theory, because it lacks physical applications.

Since the force per unit volume transforms as the components of a four-vector, we can deduce the transformation of the force on a given particle. Also, using the fact that the equation of motion of a particle is a vector equation, we can determine the transformation formula for its mass.

V. THE VARIATIONAL PRINCIPLE

In the usual theory, the complete set of field equations may be deduced using only the expressions for the field components in terms of the potential function. Half of them are obtained by eliminating the potential function from these expressions, and the other half by means of a variational principle. It will now be shown how the complete set of equations in free space ($J^2=0$) can be deduced from a variational principle, but firstly, it will be convenient to tabulate the necessary information concerning the expression of the field components in terms of the potential.

We may rewrite (6) in the form

$$-F^\nu{}_\mu = F^{\mu\nu} = T^{\mu\nu\sigma\rho} \left(\frac{\partial \kappa_\sigma}{\partial x^\rho} - \frac{\partial \kappa_\rho}{\partial x^\sigma} \right), \quad \dots\dots\dots (6a)$$

with the following condition of symmetry :

$$T^{\mu\nu\sigma\rho} = T^{\sigma\rho\mu\nu}. \quad \dots\dots\dots (6b)$$

The equation

$$F^{\mu\nu} = \frac{1}{2} T^{\mu\nu\sigma\tau} F_{\sigma\tau} \quad \dots\dots\dots (4'a)$$

is readily reduced to (4') and is therefore a consequence of (6a).

Of the scalars which it is possible to construct from $F^{\mu\nu}$ we choose the complex one

$$h = \int F^{\mu\nu} F_{\mu\nu} d\tau, \quad \dots\dots\dots (16)$$

where $d\tau$ is the four-dimensional element. It can be shown that the real part of this is equal to the corresponding scalar in the usual theory. We now vary h with respect to the potential κ_μ , keeping the geometry and coordinate system invariant to allow suffixes to be raised and lowered in varied terms. Hence

$$\begin{aligned} \delta h &= 2 \int \delta F^{\mu\nu} F_{\mu\nu} d\tau \\ &= 2 \int \left(\frac{\partial \delta \kappa_\sigma}{\partial x^\rho} - \frac{\partial \delta \kappa_\rho}{\partial x^\sigma} \right) T^{\mu\nu\sigma\rho} F_{\mu\nu} d\tau, \end{aligned}$$

by (6a)

$$= 4 \int \left(\frac{\partial \delta \kappa_\sigma}{\partial x^\rho} - \frac{\partial \delta \kappa_\rho}{\partial x^\sigma} \right) F^{\sigma\rho} d\tau,$$

by (6b) and (4'a)

$$= 8 \int \frac{\partial \delta \kappa_\sigma}{\partial x^\rho} F^{\sigma\rho} d\tau,$$

using the condition of antisymmetry. Integrating by parts, supposing that the fields vanish at infinity,

$$h = -8 \int \frac{\partial F^{\sigma\rho}}{\partial x^\rho} \partial \kappa_\sigma d\tau.$$

If \hbar is stationary for all variations in κ_μ , this requires

$$\frac{\partial}{\partial x^\rho} (F^{\sigma\rho}) = 0,$$

which are the equations (1) for free space.

VI. EARLIER COMPLEX FORMS OF MAXWELL'S EQUATIONS

The convenience of the bivectors $\mathbf{E} \pm i\mathbf{H}$ (cf. equation (5)) has already been recognized in classical electromagnetic theory (Silberstein 1907*a*, 1907*b*; Conway 1911), a reference being made by Silberstein (1907*b*) to an earlier work by Weber (1900). These bivectors, together with the quaternionic representations of the relativistic four-vectors, have been used to obtain the *Special Relativistic* form of Maxwell's equations in the form of a single quaternionic equation (Silberstein 1912, 1913) for the type of medium at present under investigation (in which $\mathbf{D}=\mathbf{E}$, $\mathbf{B}=\mathbf{H}$). In Silberstein's (1912) paper a "biquaternion" is found of which one part gives P^μ , the force and power per unit volume, while the other part corresponds to our four-vector, Q^μ . In his book, Silberstein (1924) gives the development of the quaternionic form of Maxwell's equations in classical (loc. cit., p. 46) and Special Relativity (loc. cit., p. 206) theory. However, he states (loc. cit., p. 436) that this method does not answer the purposes of General Relativity. The present work seems to indicate that the self-dual tensor (equation (4')), not the merely antisymmetrical tensor, is the appropriate generalization of the bivector. Some remarks, without references, bearing on this work are made by Sommerfeld (1948). I am indebted to the referee for pointing out the fact that the complex forms of the field vectors have been used before.

VII. REFERENCES

- CONWAY, A. W. (1911).—*Proc. R. Irish Acad.* A **29**: 1.
 SILBERSTEIN, L. (1907*a*).—Electromagnetische Grundgleichungen in bivectorieller Behandlung. *Ann. Phys. Lpz.* **22**: 579.
 SILBERSTEIN, L. (1907*b*).—Nachtrag zur Abhandlung über Electromagnetische Grundgleichungen in bivectorieller Behandlung. *Ann. Phys. Lpz.* **24**: 783.
 SILBERSTEIN, L. (1912).—Quaternionic form of Relativity. *Phil. Mag.* **23**: 790.
 SILBERSTEIN, L. (1913).—Quaternionic Relativity. *Phil. Mag.* **25**: 135.
 SILBERSTEIN, L. (1924).—"The Theory of Relativity." (Macmillan & Co.: London.)
 SOMMERFELD, A. (1948).—"Vorlesungen über theoretische Physik." Vol. 3. p. 22. (Dieterich Verlagbuchhandlung: Wiesbaden.)
 WEBER, H. (1900).—"Die partiellen Differentialgleichungen der mathematische Physik." Vol. 2. p. 348.

THE OPTIMUM SPACE-CHARGE-CONTROLLED FOCUS OF AN ELECTRON BEAM

By D. L. HOLLWAY*

[Manuscript received March 12, 1952]

Summary

The problem of space-charge defocusing of circular electron beams is considered, and expressions for the condition of optimum focus are derived from the equations of the beam profile.

It is shown that, over a wide range of spot radii, the optimum-focus expressions may be replaced by much simpler relationships suited to electron-beam design problems.

I. INTRODUCTION

The problem arises frequently in high vacuum electronics of projecting an electron beam of circular cross section to form a focused spot at some distance from an electron gun. Often the highest possible perveance is needed in the beam and, as this is increased, repulsion between neighbouring elements of charge expands the beam radially and causes a marked increase in the size of the focused spot. Usually it is not possible to introduce focusing fields after the beam leaves the gun, nor can the broadening be countered simply by over-focusing, so that in any given electrode arrangement an upper limit is set to the usable perveance.

The defocusing of convergent beams has been considered, in relation to the design of oscilloscope tubes, by Thompson and Headrick (1940), who derived expressions for the beam profiles under simplifying assumptions used earlier by Watson (1927) in the treatment of initially parallel beams.

The usefulness of these expressions is limited less by the assumptions than by the form of the equations themselves. The difficulty arises partly from the integral $\int x^2 dx$ in equation (2), and partly from the fact that the initial convergence for the condition of optimum focus is usually unknown. Thus, although the beam profile can be found for any set of initial conditions, there is still need for a simple expression relating the optimum spot size, perveance, and the electrode dimensions. The initial convergence and details of the beam profile have little influence on the main dimensions of an electrode system.

In seeking the optimum proportions of an electrode structure, it is an advantage to be able to substitute an explicit expression for the space-charge condition, to form part of a complete equation for the performance, without intermediate numerical or graphical calculations. After the geometry of the

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

system has been determined the convergence and other details of the trajectory may be calculated, or read directly from the generalized profiles shown in Figure 1 (a).

The beam profile and the condition of optimum focus will now be considered.

II. THE BEAM PROFILE

The radius r of an electron beam of circular cross section uniform in current density, after moving a distance z from the origin is given by*

$$r = R\epsilon^{-(V_z m/4eI)(V_r^2 - v_r^2)} \quad (\text{e.s.u.}), \quad \dots\dots\dots (1)$$

$$z = V_z \left(\frac{V_z m}{eI} \right)^{\frac{1}{2}} R \epsilon^{-(V_z V_r^2 m/4eI)} \int_{(V_z m/4eI)^{\frac{1}{2}} V_r}^{(V_z m/4eI)^{\frac{1}{2}} v_r} \epsilon^{x^2} dx \quad (\text{e.s.u.}), \quad \dots\dots\dots (2)$$

where R = initial radius of the beam cross section at $z=0$,

V_r = initial radial velocity of the edge electrons at $z=0$ (v_r is the corresponding velocity at distance z); these velocities are considered positive when r is increasing,

V_z = axial velocity of the beam in the direction z .

It is assumed that the surrounding space is free from fields other than that of the beam charge and that the initial radial components of velocity (i.e. the focusing velocities) are directly proportional to the radius, so that in the absence of space-charge forces the beam would converge to a point on the axis. Comparatively small edge slopes are considered, so that V_z may be written $V_z = \sqrt{2eU/m}$ (e.s.u.), where U is the beam potential.

Equations (1) and (2) are expressed conveniently in terms of the initial slope and a variable k depending only upon the perveance.

$$k^2 = mV_z^3/4eI = (e/2m)^{1/2}/(I/U^{3/2}) \quad (\text{e.s.u.}). \quad \dots\dots\dots (3)$$

(When the beam current I is measured in milliamperes and the potential U in volts, $k^2 = 0.033/(I/U^{3/2})$.)

$$\tau_1 = V_r/V_z, \quad \text{the initial slope,}$$

$$\tau = v_r/V_z.$$

Writing r_m for the minimum radius of the beam, from (1) and (2) :

$$r/R = \epsilon^{-k^2(\tau_1^2 - \tau^2)}, \quad \dots\dots\dots (4)$$

$$r_m/R = \epsilon^{-k^2\tau_1^2}, \quad \dots\dots\dots (5)$$

$$z/R = 2k \epsilon^{-k^2\tau_1^2} \int_{k\tau_1}^{k\tau} \epsilon^{x^2} dx. \quad \dots\dots\dots (6)$$

* Thompson and Headrick (1940). A misprinted "2", noted by Moss (1945), appearing in their equations (13) and (14) has been corrected in (2). Also the sign of the integral has been written to conform with the present convention for V_r . Generally similar derivations are available in texts (Spangenberg 1948; Pierce 1949).

Or, in terms of Z , the distance to the space-charge-free focus ($Z = -R\tau_1$),

$$z/Z = 2k\tau_1 \left[\Omega(k\tau_1) \pm \frac{r}{R} \Omega \left(\sqrt{\ln \left(\frac{r}{R} \right) - k^2 \tau_1^2} \right) \right], \quad \dots \dots \dots (7)$$

where

$$\Omega(y) = \epsilon^{-y^2} \int_0^y \epsilon^{x^2} dx.$$

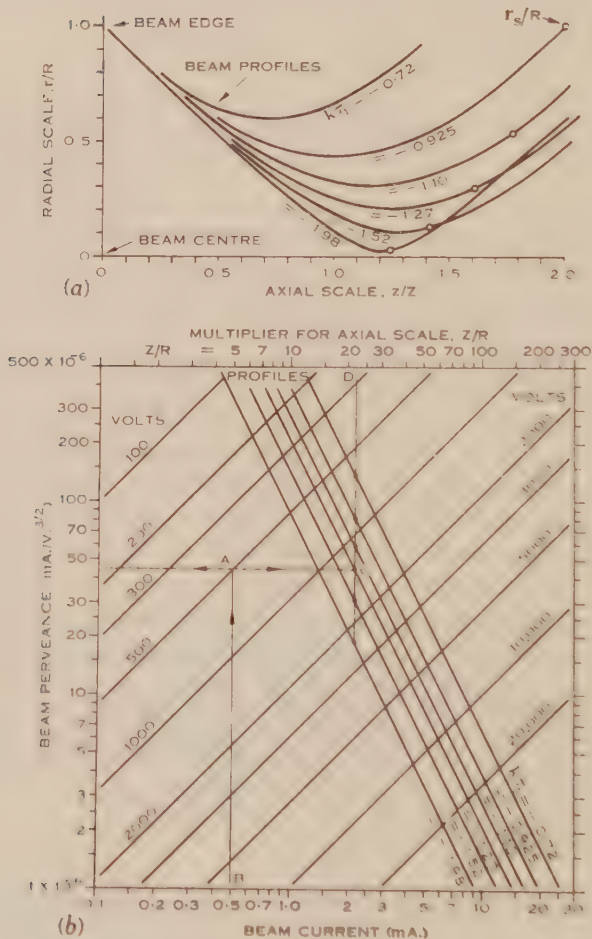


Fig. 1 (a).—Generalized profiles of beams defocused by space charge.
Fig. 1 (b).—Numerical values corresponding to Figure 1 (a).

Each (horizontal) perveance line in Figure 1 (b) makes two significant intersections, *A* with the voltage line, indicating the beam current at *B*, and *C* with a value of the parameter $k\tau_1$ denoting at *D* a multiplier Z/R . The multiplier shown ($Z/R = 21.5$) determines the axial scale of the corresponding beam profile ($k\tau_1 = -1.27$) in Figure 1 (a).

Profiles from equation (7) are drawn in Figure 1 (a), and show progressive stages in the defocusing of a beam initially directed towards the space-charge-

free focal point $z/Z=1.0$. As the perveance is raised the parameter $-k\tau_1$ decreases and the minimum radius is seen to expand, move forward, and back with respect to the space-charge-free focus. These profiles may be fitted to the numerical data of a particular problem by multiplying the axial scale of Figure 1 (a) by the value of Z/R found as shown in Figure 1 (b). As the curves are drawn for the condition of constant initial slope, perveances* corresponding to each of the six profiles are given, in Figure 1 (b), by a vertical row of intersections, for example, those shown on the line $C-D$.

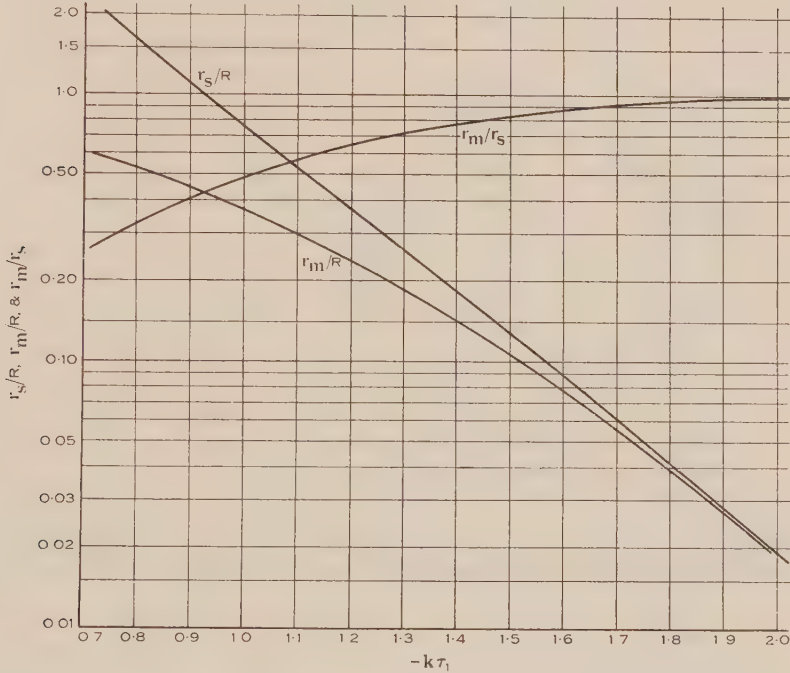


Fig. 2.—Corresponding values of r_s/R , r_m/R , and r_m/r_s are read at the three intersections of any ordinate. A convergent electron beam passes through the minimum radius r_m/R before reaching the radius of the optimum spot r_s/R .

III. THE CONDITION OF OPTIMUM FOCUS

If the profiles are to be compared for the condition of constant perveance and changes are made in the parameter $k\tau_1$ by varying τ_1 , the upper curves of Figure 1 (a) are expanded in length more than the lower by their larger multipliers ($=-1/\tau_1$). This causes the points marked (r_s/R) to fall along an envelope which denotes, at any given distance z_s , the smallest possible (or optimum) spot radius r_s .

* The beam currents corresponding to these perveances are read on the lower scale at the perveance-voltage line intersection.

The envelope satisfies the condition

$$\frac{\partial}{\partial \tau_1} \left(\frac{z}{R} \right) = 0. \quad (8)$$

Therefore, from equations (4) and (6),

$$\frac{\partial}{\partial \tau_1} \left(\frac{z}{R} \right) = 2k\varepsilon^{-k^2\tau_1^2} \left[\frac{\partial}{\partial \tau_1} \int_{k\tau_1}^{\sqrt{\ln(r/R) + k^2\tau_1^2}} \varepsilon^{x^2} dx - 2k^2\tau_1 \int_{k\tau_1}^{\sqrt{\ln(r/R) + k^2\tau_1^2}} \varepsilon^{x^2} dx \right]. \quad (9)$$

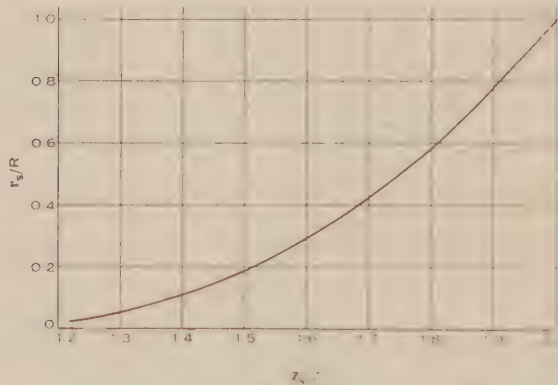


Fig. 3.—The position of the best spot z_s/R is shown as a function of the radius r_s/R .

Only the positive root is used in the upper limit as each profile passes through its minimum radius before reaching the envelope. Equation (9) reduces to

$$\varepsilon^{k^2\tau_1^2} \left[\frac{r_s}{R} \sqrt{\ln(r_s/R) + k^2\tau_1^2} - 1 \right] = 2k^2\tau_1 \int_{k\tau_1}^{\sqrt{\ln(r_s/R) + k^2\tau_1^2}} \varepsilon^{x^2} dx, \quad (10)$$

or, on substituting for r_m from equation (5),

$$2 \int_0^{\sqrt{\ln(R/r_m)}} \varepsilon^{x^2} dx - \frac{R/r_m}{\sqrt{\ln(R/r_m)}} = \frac{r_s/r_m}{\sqrt{\ln(r_s/r_m)}} - 2 \int_0^{\sqrt{\ln(r_s/r_m)}} \varepsilon^{x^2} dx. \quad (11)$$

This is of the form $F(y_1) = F(y_2)$ and it is solved by interpolation between calculated values of $F(y)$. The results, shown in Figure 2, are general and may be applied to all numerical problems. The three intersections of any ordinate representing a constant value of $k\tau_1$ denote concordant values of the minimum beam radius r_m/R , the optimum spot radius r_s/R , and r_m/r_s . The minimum is seen to approach the optimum radius asymptotically as $k\tau_1$ increases and the space-charge influence becomes less.

The axial distance to the optimum spot z_s is found also from equations (10) and (4) to (6) as follows :

$$-\tau_1 \frac{z_s}{R} = \frac{z_s}{Z} = 1 - \left(\frac{r_s}{R} \right) \frac{k/\tau_1}{\sqrt{\ln \left(\frac{r_s}{R} \right) + k^2 \tau_1^2}} \dots\dots\dots (12)$$

$$= 1 + \frac{r_s/r_m}{\sqrt{\ln (r_s/r_m)}} \cdot \frac{\sqrt{\ln (R/r_m)}}{R/r_m}, \dots\dots\dots (13)$$

or

$$\frac{z_s}{R} = k \left[\frac{1}{\sqrt{\ln (R/r_m)}} + \frac{r_s/R}{\sqrt{\ln (r_s/r_m)}} \right]. \dots\dots\dots (14)$$

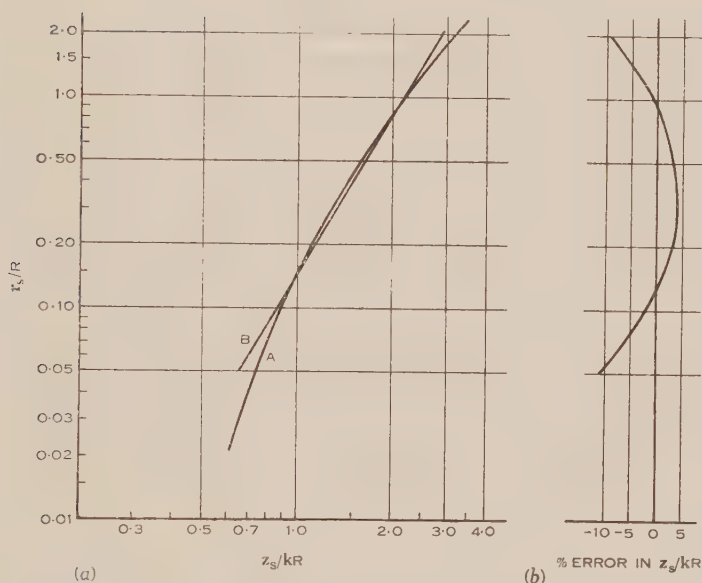


Fig. 4 (a).—*A*, the radius of the best focused spot calculated for different distances from the electron gun, compared with *B*, the approximation $r_s/R \propto (z_s/kR)^{2.5}$, used in equation (15).

Fig. 4 (b).—The deviation of equation (15) from the true value of z_s/kR .

By introducing into (13) pairs of values of r_s/r_m , R/r_m satisfying (11), the position of the best-focused spot may be compared with the space-charge-free focus for different values of r_s/R , as shown in Figure 3. When a similar substitution is made in equation (14) the bracketed term becomes a function of r_s/R alone and controls the shape of the envelope, shown at *A* in Figure 4 (a), whilst a change in the perveance displaces the whole curve axially but does not change its form.

IV. SIMPLIFICATION

The practical importance of the space-charge limitation depends to some extent on the rapid increase in the spot radius with distance from the electron gun, shown in Figure 4 (a) by curve *A*.

It has been found from the solutions of equation (14) that r_s/R is closely proportional to $(z_s/kR)^{2.5}$ (Fig. 4 (a), curve B). Thus, by combining this term with a numerical constant, it is possible to state the space-charge limitation in a single equation having the following forms:

$$\left. \begin{aligned} I/U^{3/2} &= 0.15(r_s/R)^{0.8}/(z_s/R)^2, \\ z_s/R &= 0.39(r_s/R)^{0.4}/(I/U^{3/2})^{0.5}, \\ r_s/R &= 10.5(z_s/R)^{2.5}(I/U^{3/2})^{1.25}, \\ r_s z_s &= 10.5(z_s/R)^{1.5}(I/U^{3/2})^{1.25}. \end{aligned} \right\} \dots\dots\dots (15)$$

where I = beam current in milliamperes,

U = beam potential in volts,

R = initial radius of the electron beam,

z_s = distance between the electron gun and the screen or collecting electrode,

r_s = the radius of the smallest spot which it is possible to focus at the distance z_s .

Figure 4 (b) shows that the greatest error introduced by these expressions is ~ 4 per cent. of z_s/kR in the range $0.1 < r_s/R < 1.0$. As the axial distance and the perveance have the same relationship in both the true equation (14) and in (15), the approximation applies with equal accuracy to the long low perveance beams used in oscilloscope tubes or to shorter beams of higher perveance; the error depends only upon the spot size r_s/R as shown in Figure 4 (b).

Most beam spreading calculations are approximate: the underlying assumptions, particularly that of constant beam potential, are not fulfilled exactly and the defocusing is reduced by ionization of the residual gas (Spangenberg 1948). To the accuracy needed, therefore, equation (15) may be used in the range $0.05^* < r_s/R < 2.0$.

To complete a design, the position of the space-charge-free focus may be found from the value of z_s/Z corresponding, in Figure 3, to the known spot radius r_s/R ; and this radius may be used also to select the nearest beam profile in the group shown in Figure 1 (a).

V. REFERENCES

- MOSS, H. (1945).—*J. Brit. Instn. Radio Engrs.* **5**: 10-25.
 PIERCE, J. R. (1949).—"Theory and Design of Electron Beams." pp. 147-52. (D. Van Nostrand Co.: New York.)
 SPANGENBERG, K. R. (1948).—"Vacuum Tubes." pp. 440-8. (McGraw-Hill Book Co.: New York.)
 THOMPSON, B. J., and HEADRICK, L. B. (1940).—*Proc. Inst. Radio Engrs. N.Y.* **28**: 318.
 WATSON, E. E. (1927).—*Phil. Mag.* **3**: 849.

* Other factors, including spherical aberration in the objective lens and deflexion defocusing, influence the spot size. Only the space-charge contribution is considered in equation (15).

A PRELIMINARY SURVEY OF 1420 Mc/s. LINE EMISSION FROM GALACTIC HYDROGEN

By W. N. CHRISTIANSEN* and J. V. HINDMAN*

[*Manuscript received May 14, 1952*]

Summary

A preliminary survey of 1420 Mc/s. hydrogen-line emission has been made over 270° of galactic longitude extending through the galactic centre and anticentre. The radiation source is in the form of a band of varying intensity along the galactic equator. The maximum brightness temperature is about 100°K. , in the direction of the galactic anticentre.

Measurements of line-profile show considerable variation, with a minimum width of 0.12 Mc/s. Double lines are evident between galactic longitudes 170° and 240° and it is suggested that these may originate in major structural features of the Galaxy.

The change in the peak brightness of the line along the galactic equator may result from line broadening due to galactic rotation. On the other hand it may reveal the existence of highly emitting regions. The latter interpretation is supported by the agreement in position of the bright areas for both the line emission and the continuous-spectrum galactic emission.

I. INTRODUCTION

The presence in our Galaxy of interstellar gas was indicated nearly 50 years ago, when Hartmann (1904) discovered that "stationary" absorption lines of ionized calcium were superimposed on the spectra of certain stars. Absorption lines of other atoms and molecules were also found to originate in interstellar space.

More recent work has shown that interstellar gas is distributed in the form of clouds concentrated in the galactic plane. Measurements of Doppler shift reveal that, in addition to random motions, the clouds, with the stars, revolve around the centre of the Galaxy with angular velocities which vary with the distance from the centre.

Until recently, evidence that interstellar gas consists mainly of hydrogen was only indirect, since the hydrogen was presumably in the ground state and incapable of absorbing radiation in the visible spectrum. In regions where there were very hot stars, however, faint emission lines of hydrogen were found. These lines originate from ionized hydrogen and indicate the presence of this element in the clouds.

van de Hulst (1945) suggested the possibility that line emission from ground-state hydrogen in galactic clouds might be detected in the radio-frequency part of the spectrum. A further analysis was given by Shklovskii (1949), who made a calculation of the transition probability and concluded that the line would be of sufficiently high intensity to be detected.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

The transition which is involved occurs between two hyperfine-structure sub-levels in the ground state of hydrogen. These sub-levels are produced by the effect of the nuclear magnetic moment on the electron. Laboratory measurements by Kusch and Prodel (1950) give a frequency of $1420\cdot405$ Mc s. for the radiation.

Ewen and Purcell (1951) of Harvard detected this radiation in emission from the Galaxy. They found that the radiation originated in a broad source centred on the galactic equator. The line-width, Doppler shifts, and apparent temperature of the source of the radiation were determined, and were found to be in general agreement with values expected from optical evidence, although only a small portion of the Galaxy could be studied, since a fixed aerial was used in the measurements.

Immediately following the discovery by Ewen and Purcell, Muller and Oort (1951) of Leiden also detected the line radiation. Their equipment was generally similar to that used at Harvard except for the aerial, which could be moved to view different parts of the Galaxy. Muller and Oort obtained several series of measurements at points along lines of constant declination. The source distribution at different declinations was compared, and it was found that the source followed the line of the galactic equator. Doppler shifts were measured, and the results were found to conform to Oort's model of the rotating Galaxy.

The work described in the present paper was stimulated by a letter from Professor Purcell to Dr. Pawsey of the Radiophysics Laboratory, in which the discovery of the 1420 Mc s. galactic line was communicated. It was suggested that an attempt should be made to confirm the discovery. The confirmation was announced by Pawsey (1951). Following this, those parts of the sky to which the aerial could be directed were surveyed with respect to emission of the line radiation. The results are presented here.

The investigation was exploratory, being limited by the makeshift equipment. The object was to determine the main features of the distribution in order to discover problems for future investigation and to obtain information concerning the design of new equipment.

II. EQUIPMENT

(a) *Aerial*

The aerial used for these measurements is a 25 square metre section of a paraboloid on an equatorial mount (see Plate 1). The beam width, between half-power points, for this aerial is $2\cdot3^\circ$. The side lobes are small. The aerial is normally used for solar observations at wavelengths of 25 and 50 cm. and has a range of movement in declination from $+40^\circ$ to -40° . For this work an additional aerial feed was added, and the range of declination was increased by shifting the feed in the focal plane of the reflector. This provided an increase of 10° in declination, to -50° , with little effect on the directivity of the aerial. Declination -66° was reached by a further displacement of the feed, but only at the expense of aerial directivity. The position of the aerial was indicated on hour angle and declination scales. These were checked on the Sun.

(b) Receiver

In principle, the receiver is similar to those used by Ewen and Purcell and by Muller and Oort. It is basically a superheterodyne receiver with double-frequency change. The first intermediate-frequency channel is at 30 Mc/s., with a bandwidth of 2 Mc/s. The second is at 5 Mc/s. and has a bandwidth of 0.05 Mc/s. The receiver is tuned slowly and continuously back and forth over a frequency range, which is normally 1 Mc/s., by means of the second heterodyne oscillator. When the pass-band (0.05 Mc/s. wide) of the receiver sweeps through the hydrogen line, there is a small increase in output from the receiver. This increase is so small, however, that it may be lost in receiver fluctuations; therefore, a balancing method is used to increase the sensitivity.

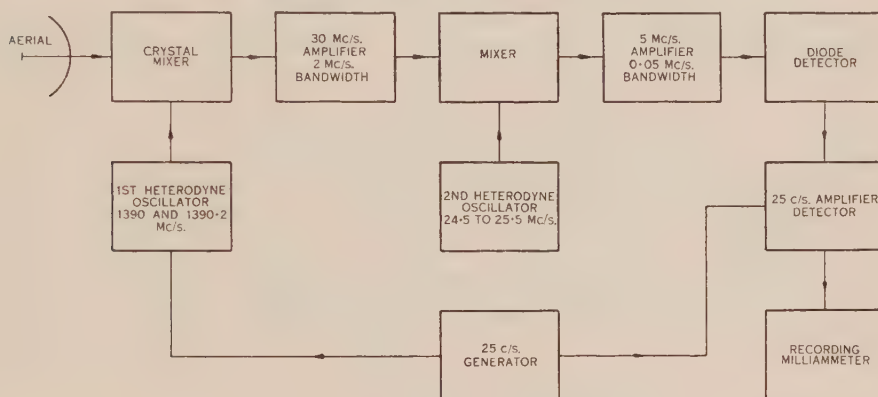


Fig. 1.—Block diagram of receiver.

The first heterodyne oscillator, which is of the line-tuned type, is switched (by means of a capacity-switch) 25 times a second between two frequencies, 0.16 Mc/s. apart, in the vicinity of 1390 Mc/s.; this causes the centre-frequency of the receiver pass-band to alternate between two positions. If the input levels at the two frequencies differ, then a 25 c/s. component will appear in the rectified output from the receiver. Such a component is recognized by means of a selective amplifier followed by a phase-sensitive detector, which is synchronized with the frequency shifting device. The general arrangement is shown in Figure 1.

As the receiver is slowly tuned over a frequency range in which a narrow spectral line is present, the energy from the line appears first in one and then in the other switched band. If the first gives rise to an in-phase 25 c/s. component as shown in Figure 2, then the output will be positive. The second will then produce an out-of-phase 25 c/s. component at the detector and the output will be negative.

The output is recorded by a recording milliammeter, in which the chart moves with known and constant speed with respect to the frequency sweep of the receiver. The chart may therefore be regarded as a record of relative intensity against frequency.

The receiver has an effective noise factor $N = 30$ when used for detection of the line radiation. The limit of detection will depend on the output fluctuations. If, as we believe true in this case, these are due to statistical noise fluctuations, these fluctuations and the receiver time constants are related in the following way. Dicke (1946) has shown that the r.m.s. fluctuation in the output expressed as a fraction of the average power output is given by

$$\frac{\Delta P}{P} = \frac{\Delta T}{(N-1)290} = \frac{\pi^{3/2}}{8} \left[\frac{B_m}{B} \right]^{1/2}, \quad \dots \dots \dots (1)$$

where B is the bandwidth of the receiver, 0.05 Mc/s., and B_m is the bandwidth of the output circuit.

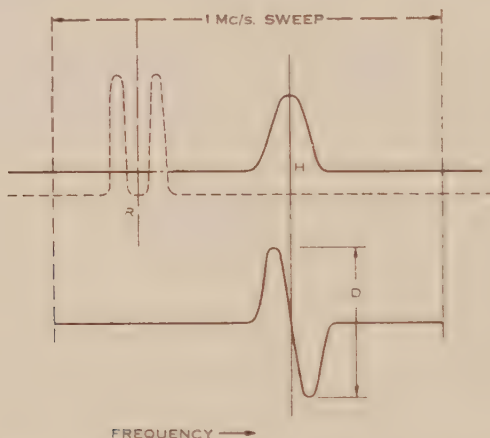


Fig. 2.—Principle of receiver operation. As the receiver pass-bands shown at R sweep across the line-profile H , the recorded output is D .

If an aerial temperature change of, say, 5° K. is the minimum change to be detected, we can tolerate an r.m.s. fluctuation ΔT of about 4.0° K. This gives B_m equal to 0.03 c/s., which means that the output of the receiver must be integrated over a period of about 30 seconds.

In practice, electrical integration over this period was not used. Instead it was found advantageous to use a shorter time constant of 6 seconds. The frequency scan was made at a rate appropriate to a 30 second time constant and the record was integrated visually.

III. METHOD

(a) *The Measurement of Line Frequency*

The second frequency-change oscillator has its frequency varied slowly by means of a rotating condenser, driven by a synchronous motor which operates from the power mains. A dial is attached to the shaft of the condenser and the position of the dial at some particular time during the rotation of the condenser is recorded on the chart of the recording milliammeter (also driven by a synchronous motor). It follows that the position of the condenser is known at

any point on the chart. The condenser dial is calibrated in terms of oscillator frequency, and is therefore an indication of frequency differences.

The absolute frequency of the received signal is determined mainly by the frequency of the first frequency-change oscillator. During the period in which the work described in this paper was done, this oscillator was of the coaxial line-tuned type with regulated power supplies. The oscillator was situated in a thermally lagged box which was fitted with a thermostatically controlled heater. Temperature variations, especially near sunrise and sunset, were responsible for variations in frequency over a range of about 0.1 Mc/s.

Line frequencies were measured by comparison with a locally injected signal of known frequency in the following way. The receiver frequency was varied at a known rate, and a record was obtained of both the hydrogen line and the slightly displaced local signal. A measurement of the relative positions on the chart of the line and local signal allowed the frequency of the former to be deduced.

The frequency of the locally injected signal was measured by comparison with a General Radio heterodyne frequency-meter, which has internal calibration by means of a crystal-controlled oscillator. The latter was checked against radio station WWV. The chief error in measurements was that involved in setting and reading the heterodyne frequency-meter. A number of measurements of the same frequency showed a standard deviation of 1 part in 20,000, i.e. about 0.07 Mc/s. in 1400 Mc/s.

(b) The Determination of Source Brightness

The excess aerial temperature due to the galactic line is equivalent to the excess brightness temperature of the line averaged over the angular width of the aerial beam. It was estimated in the following way.

The noise factor of the receiver was determined by using the Sun as a signal generator. The aerial temperature, when the aerial was directed to the Sun, was determined (for other purposes) each day, at a wavelength of 25 cm. by comparison with a thermal noise generator. The way in which the effective temperature of the Sun changes with wavelength is known fairly well (see Pawsey and Yabsley 1949). The effective temperature of the Sun at 21 cm. was taken as 0.77 of that at 25 cm. The aerial gain at 21 cm., on the other hand, was roughly $(25/21)^2$ or 1.4 times that at 25 cm. The average aerial temperature on the Sun during the months of July, August, and September 1951 was 3700 °K. at 25 cm., from which it was deduced that the value at 21 cm. was $T_s = 4000$ °K.

Measurements of the change in power output from the receiver when the aerial was directed from the sky to the Sun gave a value

$$\beta = 2.0,$$

where β is the ratio of the output powers when the aerial is directed at the Sun ($T_s = 4000$ °K.) and at the sky (where T may be taken as zero). The noise

factor N' of the receiver may be found from β , and the aerial temperature T_s and the ambient temperature T_A by use of the relation

$$(N' - 1) = \frac{T_s}{(\beta - 1)T_A}, \quad \dots\dots\dots (2)$$

giving $N' = 15$.

The effective noise factor N of the receiver, however, was twice this value, because there was no image suppression and the hydrogen line unlike the continuous spectrum of the Sun did not appear in the image band. Hence $N = 30$. With the receiver noise factor known, the noise level in the receiver was known

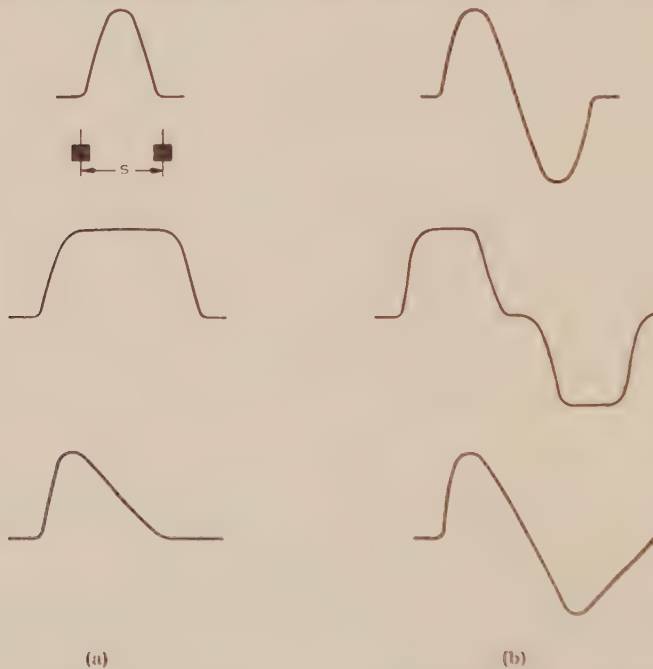


Fig. 3.—Examples of relationship between line-profile and receiver output.
(a) Arbitrary line-profiles.
(b) Corresponding receiver output.

and the fluctuations in this noise, which depend only on receiver bandwidths, could be used to calibrate the output. Using equation (1) we find the r.m.s. fluctuation in output noise expressed as a temperature to be $\Delta T = 10.7^\circ \text{K}$.

The use of the relation (1) is based on the assumption that all fluctuations were statistical noise fluctuations. If other variations were present the true temperatures will be higher than those deduced.

The r.m.s. amplitude of the noise fluctuations was determined from a statistical study of an hour's record. This was compared with the amplitude $D/2$ (Fig. 2) of the deviation in the record, caused by the hydrogen line. This provided an estimate of the brightness temperature of any particular part of the Galaxy.

(c) *The Calculation of Line-Profiles*

Since the output current from the receiver is proportional to the difference in the power received in the two switched channels of the receiver, the receiver records cannot be interpreted directly in terms of line-profile.

In Figure 3 are depicted a few arbitrary line-profiles, and next to each is the record derived from the receiver response curve. The reverse process of deriving the line-profile from the receiver record may be undertaken in the following manner.

Let $R_{(f_0)}$ be the magnitude of the receiver output when the latter is tuned to a frequency f_0 . Let the response of the receiver be represented by a function

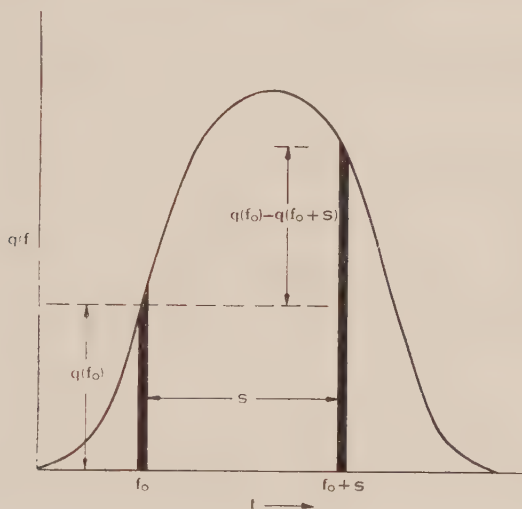


Fig. 4.—Diagram illustrating equation (4).

$p(f_0 - f)$ in one band and $p(f_0 + S - f)$ in the other. Let the line-profile be represented by a function $q(f)$. Then

$$R_{(f_0)} = k \int_{-\infty}^{\infty} q(f) \{p(f_0 - f) - p(f_0 + S - f)\} df, \quad \dots \quad (3)$$

where k is constant for a given line intensity and receiving system. In the special case where the receiver bandwidth is negligible compared with that of the line as illustrated in Figure 4, the integrand is of zero magnitude except in the immediate vicinity of f_0 and $f_0 + S$. Equation (3) then reduces to

$$R_{(f_0)} = k' [q(f_0) - q(f_0 + S)], \quad \dots \quad (4)$$

for the case where the receiver bandwidth is zero. Similarly,

$$R_{(f_0 + S)} = k' [q(f_0 + S) - q(f_0 + 2S)], \quad \dots \quad (5)$$

and so on. These equations, when added, give

$$q(f_0) = \frac{1}{k'} [R_{(f_0)} + R_{(f_0 + S)} + R_{(f_0 + 2S)} + \dots]. \quad \dots \quad (6)$$

This series may be expected to converge rapidly for most line-profiles and allows the line-profile to be calculated easily from the receiver record.

The effect of neglecting the bandwidth of the receiver is very small if the profile derived by this process has a width that is more than double the receiver bandwidth. In other cases a correction can be made by using an iterative process similar to that described by Bolton and Westfold (1950).

IV. RESULTS

(a) General

The observations described here were made between June and September 1951. The first observations were confined to the galactic centre, with the

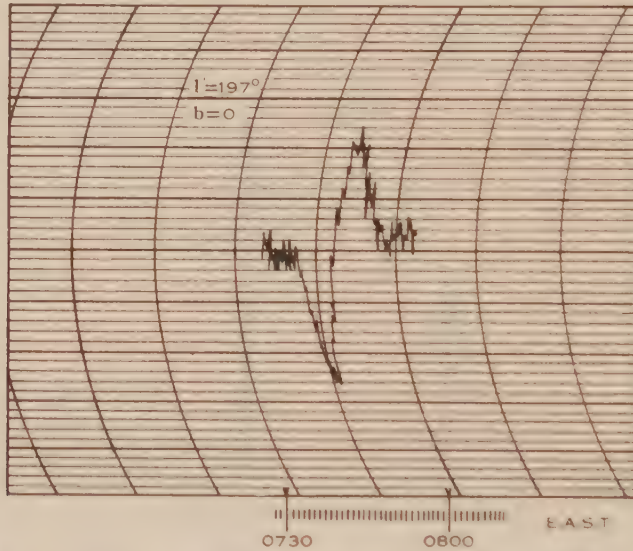


Fig. 5. —Record of line radiation in Taurus region. $l = 197^\circ$, $b = 0$.

aerial in motion to counteract the effect of the Earth's rotation. The receiver frequency was varied slowly over the full range of 1 Mc s. Later, the aerial was held stationary and the receiver allowed to scan continuously in frequency as the sky passed across the aerial beam. In this way a record was obtained every 15° in Right Ascension. This was repeated at different values of declination, so that a rough idea of the shape of the source could be obtained. Systematic measurements were then undertaken, with a change in declination of 5° each day. The range of the frequency sweep was restricted so that each sweep occupied about 20 minutes in time or 5° in Right Ascension. Near the galactic equator, measurements were taken at closer intervals by setting the aerial to follow a fixed point in the sky, and shifting it by 1 or 2° between successive measurements.

In Figure 5 is shown a record of the hydrogen-line radiation emitted from a position in the galactic equator at $l = 197^\circ$, where l is the "modified" galactic longitude centred on galactic longitude 325° . Noise fluctuations are prominent

in the record, but the expected type of response is clearly seen—a negative deflection from the mid-position followed by a sharp positive deflection, and finally a return to the normal position. The record is of a simple type and shows the effect of the “brighter” part of the Galaxy. Records from other places

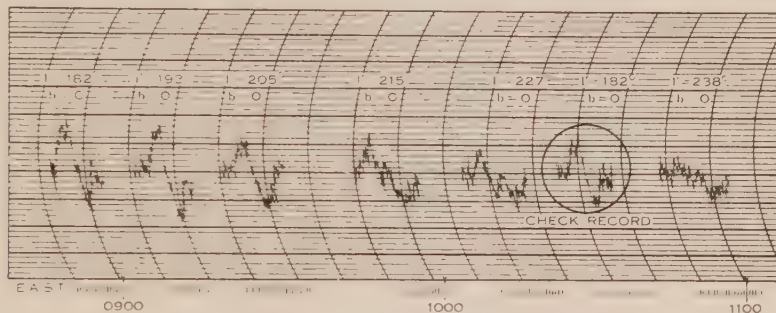


Fig. 6.—Series of records taken along the galactic equator, showing change of profile with change of longitude. (The record at $l'=182^\circ$ was repeated near the end of the observations as a check on receiver stability.)

showed smaller deflections and sometimes more complicated traces. An example of this is given in Figure 6, in which appears a series of measurements along the galactic equator. An increasingly complicated record is obtained as the aerial is moved from $l'=182^\circ$ to $l'=238^\circ$. As a check the measurement at $l'=182^\circ$ was repeated near the end of the series.

(b) Brightness Distribution

The maximum deviation in the receiver output, shown as D in Figure 2, was taken as a measure of the peak brightness in different parts of the sky.

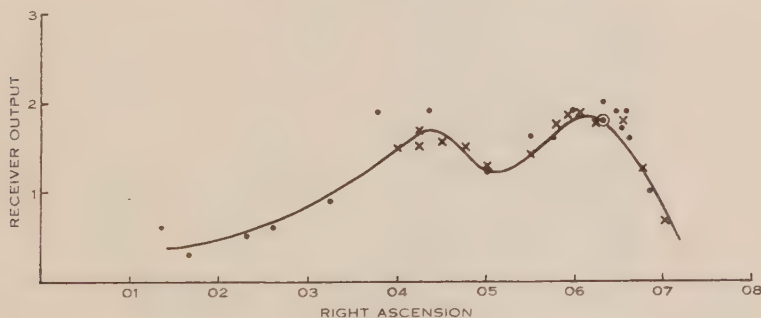


Fig. 7.—Measurements of peak brightness at Declination $+10^\circ$.

This will be correct, provided the line-width is appreciably greater than the receiver bandwidth, but less than the spacing between the two receiver channels. The line-profile indicated that over a large part of the sky this condition was fulfilled. However, some values of brightness will be lower than the true ones.

Records were obtained at roughly 5° intervals in Right Ascension over 24 hours, with a fixed value of declination, which was changed each day. Several

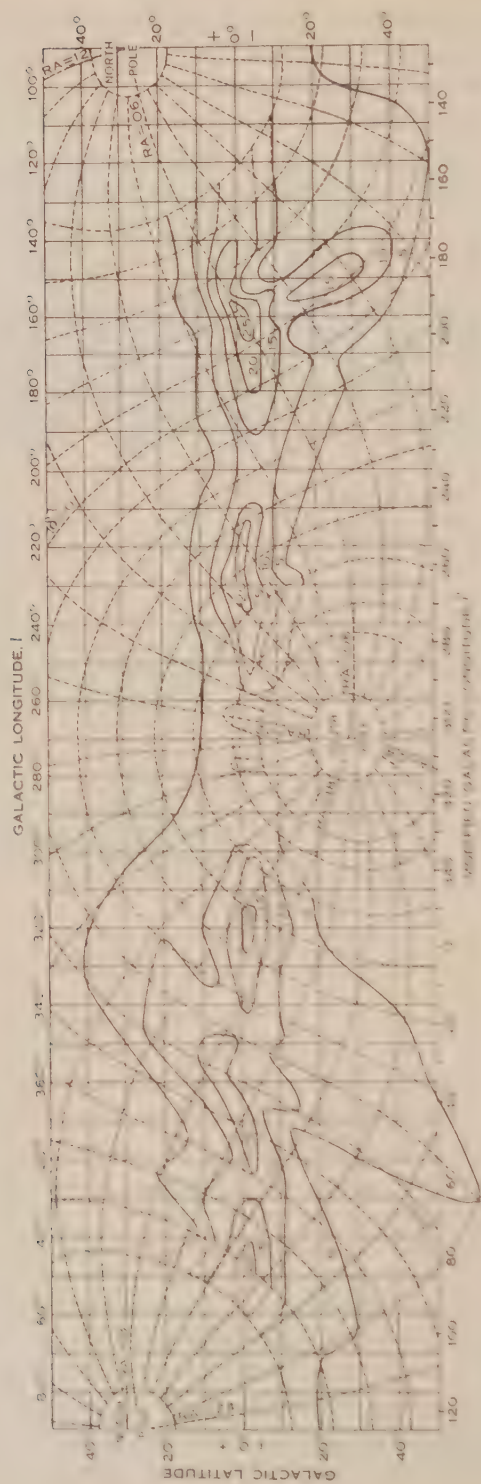


Fig. 8. 1420 Mc γ hydrogen line radiation. Galactic chart giving contours of peak brightness. Twenty-five units corresponds to a brightness temperature of about 100 $^{\circ}$ K.

complete sets of measurements were made at each declination and it was found that differences in magnitude of up to 20 per cent. existed between different sets of readings. The scatter at one declination is shown in Figure 7, individual sets of measurements being indicated by different symbols. A curve of line brightness against position in Right Ascension may be drawn for each fixed value

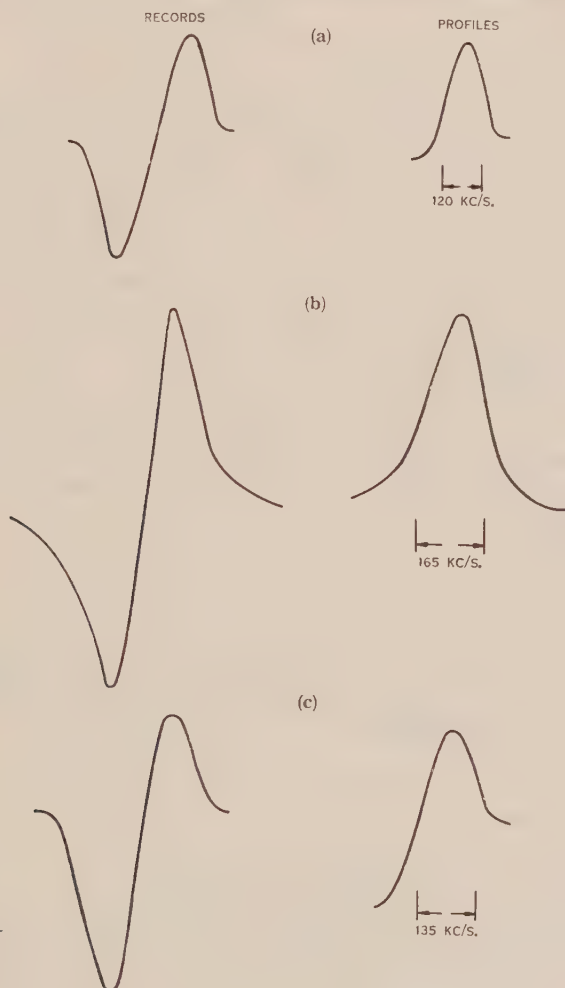


Fig. 9.—Some smoothed records and line-profiles.

(a) Galactic centre $l'=0$, $b=0$.

(b) Anticentre region $l'=195^\circ$, $b=0$.

(c) Cygnus region $l'=80^\circ$, $b=0$.

of declination. Another series of measurements at points where the lines of fixed declination cross the galactic equator provides a common scale for the different curves and allows contours of peak line brightness to be drawn for those parts of the sky to which the aerial could be directed. The quantities marked in the diagrams are in arbitrary units. The maximum corresponds to a

temperature of about 100°K . The contours are shown in Figure 8. The estimated relative error in the measurements on which the curves are based is 15 per cent. In some parts of the map, the position of the outer contour could not be fixed accurately because the intensity gradient was small and the intensity was close to the minimum level detectable with this equipment.



Fig. 10.—Smoothed record and double line-profile $l' = 55^\circ$, $b = 0$.

The dotted contours are based on measurements which involve greater error than that quoted above.

(c) Line-Profiles and Centre-Frequencies

Figure 9 shows smoothed records of the line for three positions in the Galaxy and an estimate of the line-profile at each position: the line-width varies between 0.12 and 0.16 Mc/s.

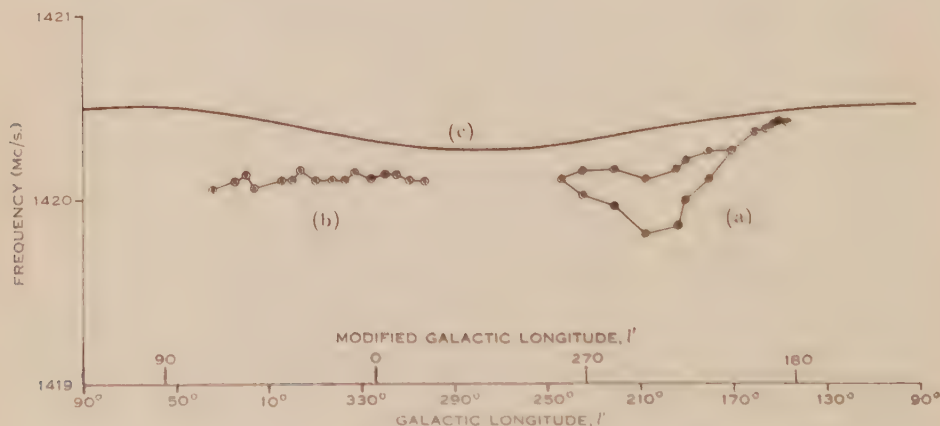


Fig. 11.—Line centre-frequencies measured along the galactic equator.

- (a) Measured line centre-frequency showing effect of double line.
- (b) Measured line centre-frequency (double lines were detected in this region but measurements were incomplete).
- (c) Frequency variation due to assumed Earth's motion relative to local stars.

In Figure 10 (a) a different kind of trace is shown: this has the appearance of a double line. Such traces were obtained over a considerable range of angle along the galactic equator. In Figure 10 (b) an estimated profile of the line is shown. The lines are separated by 0.23 Mc/s, and each has a width of about 0.13 Mc/s.

The centre-frequencies of the line or lines were determined for points along the galactic equator in the manner described earlier. These frequencies are plotted in Figure 11. In the region marked (*a*), the double line persists over a large range of galactic longitude. In region (*b*) a double line was also seen, with separation similar to that in (*a*). Measurements were incomplete, however, and the results have not been plotted.

V. DISCUSSION OF RESULTS

(*a*) Confirmation of Existence of Galactic Hydrogen

The most obvious feature of the measurements is that the 1420 Mc/s. hydrogen line originates in a source which forms a band of varying angular width, centred on the galactic equator. Hence it follows that ground-state hydrogen exists in space, and its distribution in the Galaxy is broadly similar to that of the visible stars, i.e. it is concentrated in the equatorial regions.

(*b*) Causes of the Difference in Brightness

The second feature of the measurements is the variation in brightness along the galactic equator. Two factors may contribute to this effect, galactic rotation and broad structural features of the Galaxy. The effect of galactic rotation will be considered first.

It is believed that the Galaxy rotates about its axis in such a way that the angular velocity varies with the distance from the centre. This non-uniform rotation produces relative motions between different parts of the Galaxy. The velocities of the various parts of the Galaxy relative to the Sun, have, in general, a radial component which varies both with the distance from the Sun and with direction. The effect in direction is that the radial velocities vary sinusoidally through two periods as the direction changes through 360° around the galactic equator. Zeros are found in the four cardinal directions referred to the galactic centre.

The variation in radial velocity with distance would be expected to produce a broadening of spectral lines. In an optically thin medium the energy of the line would be spread over a range of frequencies, and the peak brightness of the line, as defined earlier, would be reduced. In an optically thick medium, the peak brightness would be unaffected. Since the broadening effect would change with galactic longitude, a periodic change in line-width and possibly in peak brightness would be expected, if the medium is approximately uniform.

A rough quantitative estimate of the effect can be derived as follows. It is assumed that the hydrogen is uniformly distributed in the plane of the Galaxy, and that this plane extends indefinitely. It is also assumed for ease of calculation that the internal turbulent movements in the galactic hydrogen are such that there is an equal probability that the gas, at any particular distance from the observer, has a radial velocity with respect to the Earth lying somewhere in the range v to $(v+\Delta v)$. In the absence of galactic rotation, the range of radial velocities will be Δv and the corresponding line-width Δf related by

$$\frac{\Delta v}{c} = \frac{\Delta f}{f}, \quad \dots\dots\dots (7)$$

where f is the frequency of the line radiation and c is the velocity of light. To calculate the effect of rotation, one can use a formula derived by Oort for use when the regions involved are within a few thousand parsecs of the Sun. This is

$$v = r.A. \sin 2l', \quad \dots\dots\dots (8)$$

where v is the radial velocity resulting from galactic rotation,

r is the distance of the source from the Sun,

A is 6×10^{-16} sec.⁻¹,

l' is the "modified" galactic longitude measured with respect to the galactic centre.

In order to calculate the optical thickness of the medium one may consider the absorption in interstellar space of monochromatic radiation leaving the Earth. In directions where the radial velocity, produced by galactic rotation, is changing with distance, the radiation will be absorbed over only a restricted range of

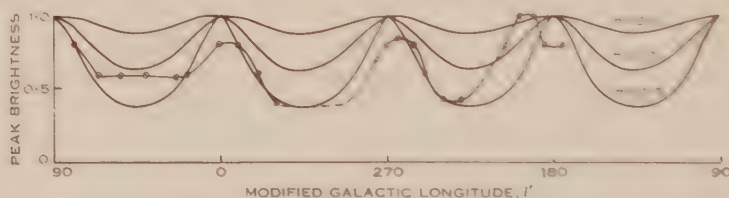


Fig. 12.—Effect of galactic rotation on peak brightness.

— Calculated effect for a uniform medium.
 ⊙—⊙ Measured values.

distance Δr . The magnitude of Δr depends on the internal motion of the medium and falls to zero when there is no internal motion. If the range of radial distances Δr is chosen so that the range of radial velocities resulting from galactic rotation is equal to the range of velocities Δv of internal motion, given by equation (7), then radiation of the chosen frequency will be absorbed equally at all points in the range Δr , and not at all outside this range. Δr is related to Δv by use of equation (8) which gives

$$\Delta r = \Delta v A \sin 2l'. \quad \dots\dots\dots (9)$$

Combining equations (7) and (9) we obtain

$$\Delta r = \frac{c.\Delta f}{f.A.\sin 2l'} \quad \dots\dots\dots (10)$$

The optical thickness of the medium may now be calculated if the absorption coefficient K_a is known. If the expression for K_a used by Wild (1952) is employed, we find

$$K_a = \frac{8.0 \times 10n}{\Delta f.\theta} \text{ parsec}^{-1} = \frac{2.6 \times 10^{-15}n}{\Delta f.\theta} \text{ cm.}^{-1}, \quad \dots\dots (11)$$

which gives

$$\tau = K_a.\Delta r = \frac{2.6 \times 10^{-15}n.c.\Delta f}{\Delta f.\theta.f.A.\sin 2l'}, \quad \dots\dots\dots (12)$$

where n is the number of ground-state hydrogen atoms per cm.³ in the medium and θ is the excitation temperature of the gas.

If θ is taken as 100°K. , and the relative brightness of the source is plotted for various values of l' and n , the family of curves shown in Figure 12 is obtained. Measured values along the galactic equator are also plotted.

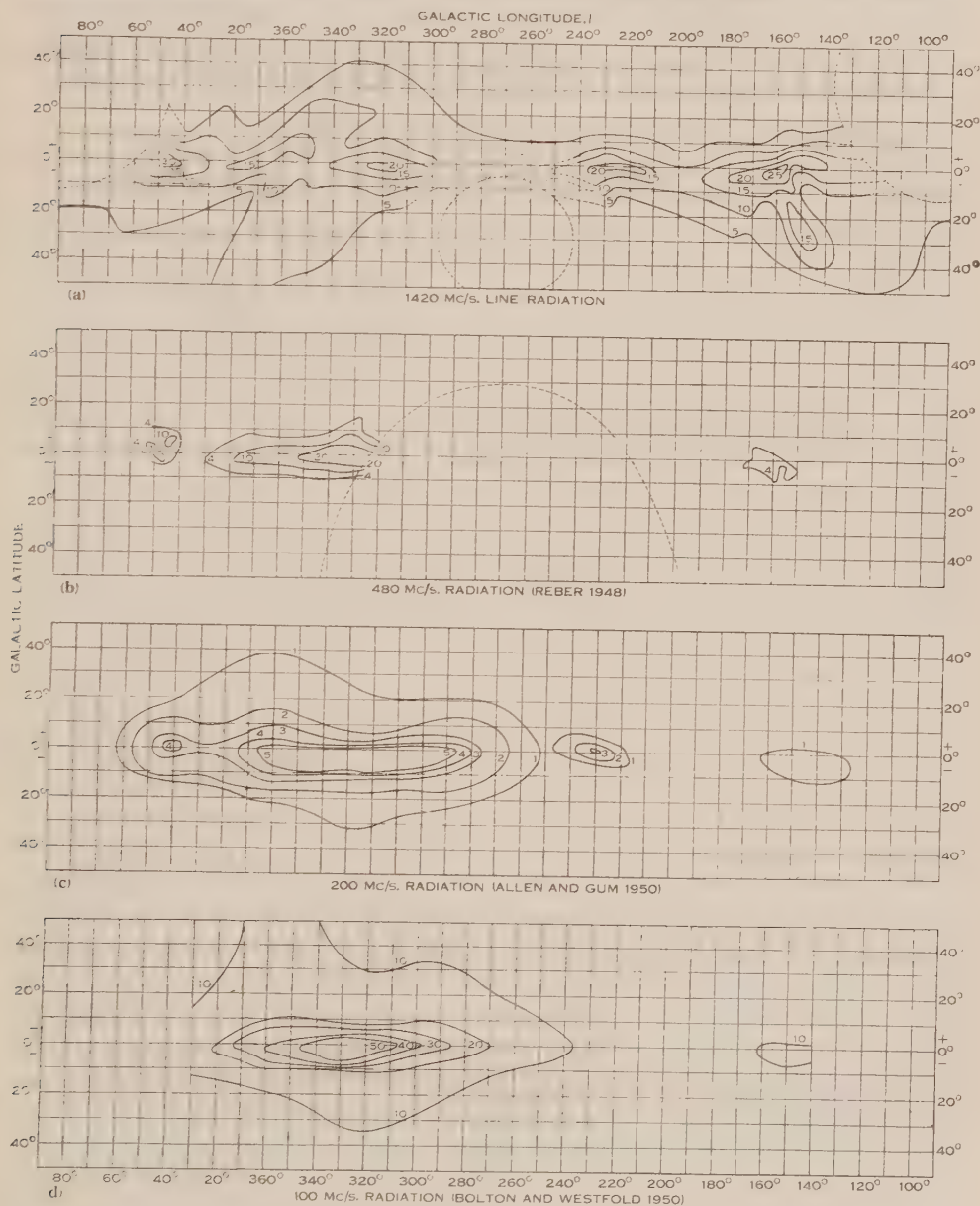


Fig. 13.—Brightness distribution of galactic radio-frequency radiation.

It will be seen that there is a very rough agreement between the measured values and those deduced from the crude model, if n lies somewhere between 0.5 and 1.0 atoms per cm^3 . The value of n estimated from optical evidence

is in agreement with this. It appears from the evidence on line-widths, however, that the assumed model is inadequate. There is no appearance of a continuous broadening of the line in directions near $l' = 45^\circ, 225^\circ$; instead, double lines appear, each having a width about equal to that of the single lines in the directions $l' = 0, l' = 180^\circ$. In addition there are irregularities in the contours that cannot be explained in terms of galactic rotation. Galactic rotation is capable of explaining only part of the variation in brightness of the source.

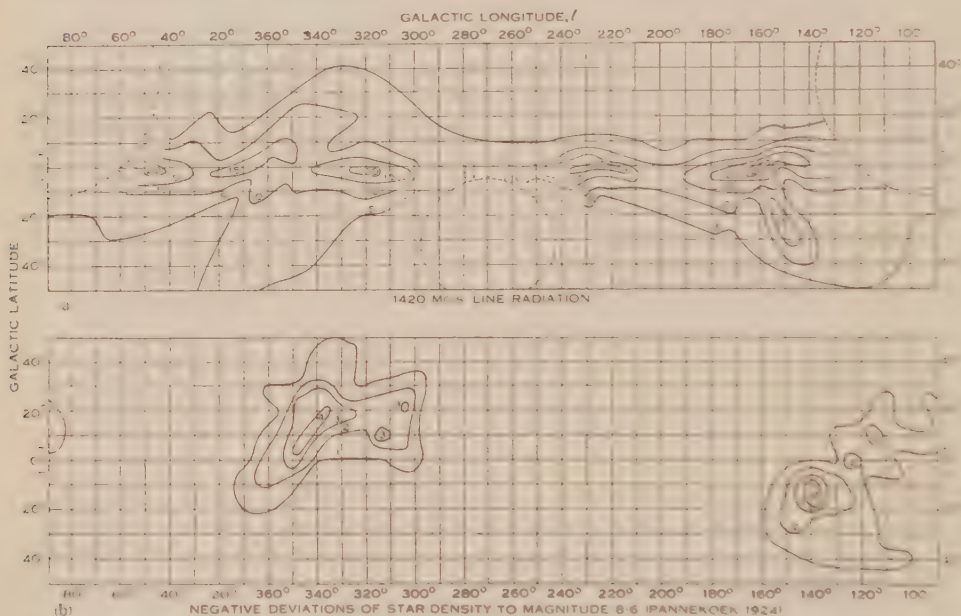


Fig. 14.—Comparison of contours of line brightness and local interstellar obscuration.

If the differences in line brightness represent inhomogeneities in the Galaxy, it should be possible to find some correspondence with other optical or radio features. In Figure 13 are shown brightness contour maps of the continuous galactic radio emission at various frequencies and, for comparison, a map on the same scale of the 1420 Mc/s. line emission. There is a close correspondence between the positions of the maxima in the continuous and line emissions. The continuous radiation will not be affected appreciably by galactic rotation; hence the correspondence of the contours strongly indicates that the common characteristics are due to structural features of the Galaxy which affect both types of radiation.

The principal differences between the distribution of the line and continuous radiation are, firstly, the concentration towards the galactic centre in the case of the continuous radiation, and secondly, the extensions in the Taurus and Ophiucus regions in the case of the line radiation. The first could be the result of greater penetration of the continuous radiation compared with the line

radiation. The extensions in the Taurus and Ophiucus regions may be related to the two obscuring cloud systems (Fig. 14), discovered by Pannekoek (1924) from studies of negative deviations in star counts up to magnitude 8.6. Pannekoek estimated that these clouds were at a distance of roughly 100 parsecs from the solar system. If these clouds are related to the hydrogen-line distribution then one may infer that dust and hydrogen are associated in them.

(c) *The Significance of the Double Lines*

The existence of double lines of the type shown in Figure 10 indicates the presence of two sources having different radial velocities. The frequencies and angular extent of these lines are shown in Figure 11 (a), and their appearance, particularly the roughly sinusoidal variation of line separation over a quadrant, suggests that the radial velocities result from galactic rotation. If this is so, the approximate relation between radial velocity and distance given in equation (8) may be used to determine the distances to the two sources. A correction has first to be made for the Sun's motion relative to local stars. This is shown in Figure 11 (c). If the value 1420.4 Mc/s. is taken as the natural line frequency, the measured line frequencies are low by about 0.1 Mc/s. (The directions where the effect of galactic rotation is zero are significant in this respect.) Since the difference is inside the limit of error, it has been assumed to be a systematic error, and a correction has been made for it. The distance to the two sources, as estimated from equation (8), is then 1000 and 4000 parsecs. These distances are fairly constant over half the quadrant. If this conclusion is correct, the vast size and approximately constant separation of the two hydrogen sources suggest that they are in the form of spiral arms of the Galaxy.

(d) *Estimate of Thickness of Galaxy*

If the distances derived above are approximately correct, we can use them to estimate d , the half-thickness of the hydrogen perpendicular to the galactic plane. In the direction of the anticentre, the distance of the remote source is 4000 parsecs and the angle which it subtends is approximately 8° . This gives $d=250$ parsecs.

As shown by Wild (1952), the brightness distribution of the source with respect to galactic latitude allows an estimate to be made of the product of galactic thickness (perpendicular to the galactic plane) and the absorption coefficient K_a . In the direction of the galactic centre the product $K_a d$ is equal to 0.15. If in the expression (11) for K_a , we put the line-width, $\Delta f = 0.12 \times 10^6$ c/s. and $\theta = 100^\circ \text{K.}$, we find that nd is approximately 200. If we assume $n = 0.7$ atoms per cm.^3 we find the galactic half-thickness d is about 300 parsecs. These two values for d , 250 and 300 parsecs, are much greater than the value 50 parsecs estimated by Muller and Oort.

(e) *Motion of the Source*

Evidence of the motion of the source of 1420 Mc/s. radiation relative to the Earth is based on line-profiles and centre-frequencies. In the observations described here this information is limited; much more complete information should be obtainable in the future.

Random motions of the sources may be determined from line-widths in directions where galactic rotation has no effect. Measured widths were found to lie between 0.12 and 0.17 Mc/s. It may be assumed that these widths result from local mass movements in the Galaxy rather than temperatures of the order of 10^4 °K. The corresponding random radial velocities are about ± 12 to 18 km. per sec. These values are greater than those quoted by Ewen and Purcell (1951) and Muller and Oort (1951). It seems possible that the smaller value for the line-width obtained by Ewen and Purcell resulted from the narrow spacing between the two reception bands of their receiver. Regarding the value of Muller and Oort we believe*† that their unpublished measurements of line-width are in agreement with ours. The effects of galactic rotation on line frequency and profile are not clear cut, due presumably to the inhomogeneous nature of the gas distribution (Figs. 11 and 12). Evidence of effects due to the Sun's motion relative to local star groups is similarly limited.

(f) *Other Physical Properties*

Measurements of the brightness temperature in the direction of the galactic centre, where the optical thickness is certainly great, and at the anticentre are in close agreement at about 100 °K. if allowance is made for the temperature of the continuous radiation (about 15 °K.) around the galactic centre. In these directions the brightness temperature should be close to the excitation temperature for the line. It seems probable that this in turn is close to the kinetic temperature of the gas.

Earlier in the discussion some evidence was produced (see Fig. 12), which was consistent with an atom density between 0.5 and 1.0 atoms per cm.³ in the Galaxy. This is about the value that has been deduced from optical information. With such a density and with values of line-width and excitation temperature, quoted above, the distance to optical depth unity, in the absence of effects of galactic rotation, may be calculated from equation (11) to be about 2000 parsecs. This is much greater than the values derived by Ewen and Purcell, and Muller and Oort. As pointed out by Muller and Oort, galactic rotation will allow penetration to greater distances. In at least one quadrant there is evidence, given above, that penetration to a depth of 4000 parsecs is possible.

VI. CONCLUSIONS

The investigation described here has shown clearly that a detailed and precise study of the line radiation from galactic hydrogen will yield most important astronomical information. The preliminary work, however, has led to certain definite conclusions. The first is that the existence of galactic hydrogen is confirmed. The observations showed that the source of the line radiation

* Based on sketches given by Professor Oort to Mr. F. J. Kerr.

† *Note added in Proof.* Dr. H. C. van de Hulst has since informed us that the measurements in Holland indicate a bandwidth of about 136 kc/s. at the galactic anticentre. In most directions the line is wider.

LINE EMISSION FROM GALACTIC HYDROGEN



Twenty-five square metre aerial and polar mount.

occupies roughly the same part of the sky as does the visible Milky Way. Hence it may be assumed that the hydrogen is concentrated near the equatorial plane of the Galaxy.

The appearance of double lines over a considerable range of galactic longitudes is an important observation. It suggests the possible existence of spiral arms in the Galaxy. A more thorough study of line-profiles should provide definite conclusions with regard to this.

Another feature of the observations is the differences in line brightness that occur along the galactic equator. Whether these variations indicate structural features of the Galaxy, as is suggested by partly similar variations in the continuous spectrum of galactic radiation, or are the effect of galactic rotation, is a question that is not yet answered. The maximum brightness temperature observed was at the anticentre, and was about 100 °K.

The measurements of line-widths in certain directions corresponded to random radial velocities between 12 and 18 km/s.

A tentative estimate was made of the thickness of the hydrogen source, in the direction perpendicular to the galactic plane. The density of hydrogen atoms in the Galaxy was also estimated. Definite conclusions on these and other questions require measurements of the whole frequency profile of the line, obtained with equipment of adequate frequency stability and of the utmost sensitivity.

VII. ACKNOWLEDGMENTS

The work reported here was marked by close contact between groups of workers in different countries. The writers wish particularly to thank Professor E. M. Purcell and Dr. H. I. Ewen for their generosity in supplying pre-publication information on the work at Harvard, and Professor J. H. Oort, Dr. H. C. van de Hulst, and Mr. C. A. Muller of Leiden for their ready cooperation. Mr. F. J. Kerr acted as a link between the three groups and provided stimulating comment on the work. Mr. J. G. Bolton and Mr. J. P. Wild also helped the authors in useful discussion. We are also indebted to Dr. J. L. Pawsey, whose enthusiasm and criticism were greatly valued. We wish to acknowledge gratefully the assistance given in observations or in the preparation of equipment by Messrs. J. Warburton, A. J. Harragon, G. Fairweather, G. J. Stanley, and G. H. Trensky.

VIII. REFERENCES

- ALLEN, C. W., and GUM, C. S. (1950).—*Aust. J. Sci. Res. A* **3**: 224-33.
 BOLTON, J. G., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res. A* **3**: 19-33.
 DICKE, R. H. (1946).—*Rev. Sci. Instrum.* **17**: 268.
 EWEN, H. I., and PURCELL, E. M. (1951).—*Nature* **168**: 356.
 HARTMANN, J. (1904).—*Astrophys. J.* **19**: 263-86.
 VAN DE HULST, H. C. (1945).—*Ned. tijdschr. Naturkunde* **11**: 201.
 KUSCH, P., and PRODELL, A. G. (1950).—*Phys. Rev.* **79**: 1009.
 MULLER, C. A., and OORT, J. H. (1951).—*Nature* **168**: 357-8.
 PANNEKOEK, A. (1924).—*Publ. Astr. Inst. Univ. Amst.* No. 1.
 PAWSEY, J. L. (1951).—*Nature* **168**: 358.
 PAWSEY, J. L., and YABSLEY, D. E. (1949).—*Aust. J. Sci. Res. A* **2**: 198-213.
 REBER, G. (1948).—*Proc. Inst. Radio Engrs. N.Y.* **36**: 1215-18.
 SHKLOVSKII, I. S. (1949).—*Astr. Zh.* **26**: 10.
 WILD, J. P. (1952).—*Astrophys. J.* **115**: 206-21.

THE POSITIONS OF SIX DISCRETE SOURCES OF COSMIC RADIO RADIATION

By B. Y. MILLS*

[*Manuscript received March 5, 1952*]

Summary

The positions of six discrete sources of cosmic radio radiation have been measured accurately. The tentative identifications of three of the sources with nebulae made by Bolton, Stanley, and Slee appear to be confirmed. It is also found that two of the sources have measurable angular sizes (about $\frac{1}{2}^\circ$).

I. INTRODUCTION

In a previous paper (Mills 1952) the statistical distribution of 77 discrete sources was examined. In obtaining their positions, however, the full accuracy of the interferometer was not utilized. Further observations of six of these sources have now been undertaken to find their positions with the greatest accuracy attainable with the equipment, which is a few seconds of time in Right Ascension and a few minutes of arc in declination. The sources observed are Cygnus-A (19+4), Centaurus-A (13-4), Virgo-A (12+1), Taurus-A (05+2), Hydra-A (09-1A), and Fornax-A (03-3), where the names of the sources are those used by Stanley and Slee (1950) and the numbers are the reference numbers used by Mills (1952).

The positions of Taurus-A, Centaurus-A, and Virgo-A fall close to those given by Bolton, Stanley, and Slee (1949), and serve to strengthen the tentative identifications which they made with the nebulae M 1 (the Crab), NGC 5128, and NGC 4486 respectively. The position of Cygnus-A falls close to that obtained previously by Mills and Thomas (1951). There is no outstanding stellar object near this position, nor is there any in the vicinity of the Hydra-A and Fornax-A sources. Recently Smith (1951) has published positions of the Cygnus, Taurus, and Virgo sources which agree well with the present ones. It may therefore be said that the positions of these three sources and that of Centaurus-A have now been established with some degree of certainty.

It is interesting that the angular sizes of two of the sources studied (13-4 and 03-3) are greater than the limit of resolution of the equipment, which is about 10' of arc. The equivalent width of each in the east-west direction is about 20' of arc.

II. EQUIPMENT AND MEASUREMENTS

The receiving equipment has already been described (Mills 1952); briefly it consists of three 100 Mc/s. aeriads in the form of broadside arrays, placed on

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

an east-west line with a spacing between the outside aerials of 900 ft. and between the closer inner pair of 200 ft. The outputs of these aerials can be separately combined in pairs to form Michelson-type interferometers. Two receivers are normally employed, operating with the largest and smallest available spacings (900 and 200 ft.). A 25 c/s. lobe switching system is used for isolating the discrete sources from the noise background.

In operation the electrical lengths from the signal mixing points to each aerial are fixed at known values by a calibrating system which is similar to that described by Mills and Thomas (1951) and Little and Payne-Scott (1951). The transit time of the source is then related to the occurrence of zeros in the sinusoidal interference patterns produced on the pen recorders. However, it is not possible to decide immediately which is the zero corresponding to transit of the source, and three steps are required to determine the Right Ascension. Firstly, an approximate value is obtained by noting the sidereal time at which the interference pattern has its maximum amplitude. An improvement in accuracy is then obtained by noting the times of zeros in the pattern produced by the close-spaced pair of aerials. The position so obtained is sufficiently accurate to identify the zero of the wide-spaced aerial pattern corresponding to transit of the source, which then gives the source position with much greater accuracy, but still subject to some corrections which will be discussed later.

The declination is obtained by noting the times of occurrence of 20 zeros on each side of the central one, and calculating in the way described previously (Mills and Thomas 1951). As before, the bandwidth is restricted to 100 kc/s. and a crystal-controlled local oscillator is used in order to define the frequency accurately.

The angular size is estimated by measuring the apparent intensities of a source when using the close-spaced and wide-spaced aerials. If the source extends in an east-west direction the intensities will be different, and their ratio can be used to calculate an equivalent angular size in the manner of McCready, Pawsey, and Payne-Scott (1947). The smallest extension which can be measured in this way is about $10'$ of arc.

III. ACCURACY

The errors which can occur in a Michelson-type interferometer have been discussed previously (Mills and Thomas 1951). Similar errors occur in the present equipment but their relative importances are different. The errors are of two types, systematic and random, which will be considered in turn.

(a) *Systematic Errors*

Previously, differences in the ground reflection effects at the two aerials had been the most important source of errors in both the Right Ascension and declination measurements; now the errors are much reduced for two reasons:

- (i) The aerials are much more directive in the vertical plane, so that the ground-reflected ray is of smaller intensity and has less effect on the phase of the combined signal.
- (ii) The ground is much flatter in the vicinity of the aerials, so that the phases of the reflected rays at each aerial are much more nearly equal.

The only correction required was one of $+1$ second in the Right Ascension of the Cygnus source.

The most serious source of equipment error now lies in the lack of precise knowledge of the relative positions of the electrical centres of each of the aerials. Previously the difficulty had been overcome by interchanging aerials, but now the aerials are too large and reliance has been placed instead on making the aerials as similar as possible and afterwards checking by careful electrical measurements. The measurements involved, firstly, the careful matching of one aerial and, secondly, the measurements of the matching achieved with the other two aerials after their transformers had been set identically with that of the first aerial. Also various parts of the feed and aerial system were short-circuited in turn, and impedances measured at the point of connection of the feed cables with the calibrating switches. Where necessary, adjustments were made in the length of feed cables so that all aerials displayed identical impedances under any of these tests.

The probable error in the relative locations of the electrical centres of a pair of aerials due to their dissimilarity has been estimated as 1 in., resulting in a probable error in Right Ascension of $\frac{1}{3}'$ of arc, and in declination of $\frac{1}{2}'$ of arc at a declination of 40° (i.e. a probable error equivalent to an error in the measured period of a pattern of 1 part in 10,000). Surveying errors, for which it is estimated that the probable error is less than a quarter of an inch, are negligible by comparison.

Two further equipment errors can occur in the Right Ascension measurement due to the calibrating system and to time delays in the recorder and its associated amplifying system. Errors in the calibrating system were found to be very small in the previous work. An improvement in the switching system has now been made, and it is considered that systematic errors due to this cause can be ignored. Time delays in the recorder and amplifying equipment are important, particularly with the weaker sources when integrating time constants up to 8 seconds were used. For calibration, the recorder response curve was obtained to a step function applied right through the equipment under operating conditions. The delay time produced on the normal sinusoidal signal was then calculated from this curve. For 19 ± 1 (time constant 1 second), 0.5 ± 2 , 1.3 ± 4 , and 12 ± 1 (time constant 4 seconds) the errors are negligible. For $0.9 \pm 1A$ and 0.9 ± 3 (time constant 8 seconds), the error may amount to 1 second.

Finally, the declination is subject to a systematic equipment error due to the uncertainty in operating frequency. It is estimated that the probable error due to this cause is equivalent to an error in the measured period of a pattern of about 1 part in 10,000.

There are also two external sources of systematic error. Firstly, refraction, which can cause an error in the declination, has been discussed before but the corrections have now been calculated in a slightly different way that is a little more accurate. It can be shown that for an interferometer with a plane atmosphere the refraction correction is automatically applied if the free-space velocity of radio waves is used in the declination calculation. The effect of taking account of the Earth's curvature is negligible for the troposphere, and the only

remaining correction is that due to the curvature of the ionosphere. This is obtained by calculating the ionospheric refraction correction in the same way as before. Since the total correction to be applied is now for the ionospheric refraction alone and is much smaller than before, the result is more accurate.

The second external source of error which is likely to be particularly bad on the smaller sources is due to a displacement of the zeros of the interference pattern caused by adjacent sources also in the aerial beam. Errors of this kind may be minimized by averaging the times of a number of zeros. Normally the Right Ascension is obtained from the times of 14 zeros, and the declination from the times of 40 zeros. Since the errors to be expected from this cause may be estimated from the modulation observed on the interference patterns, they are discussed later when dealing with the individual sources.

Summing up, the important cause of systematic error in the Right Ascension measurement is due to uncertainty in the location of the electrical centres of the aerials, and results in a probable error estimated to be $\pm\frac{1}{3}'$ of arc. There are several systematic errors in the declination measurement of nearly equal importance. It is estimated that their joint contribution represents a probable error in the measured period of 1 part in 4000, that is, about $\pm 1'$ of arc at a declination of 40° . In addition to these basic errors which are the same for all sources, there are further uncertainties due to interference from nearby sources, for which allowance must be made individually.

(b) *Random Errors*

The only random errors which were of any importance were due to errors in the calibration and to refraction effects. Random errors can occur in the calibration system owing to inaccurate setting and to changes in the calibration over a period of days (normally the calibration was carried out only two or three times a week as the equipment has proved considerably more stable than before). The random refraction effects appear to be less than those observed on the Cygnus source previously both as regards apparent position variations and intensity fluctuations. The probable error due to random effects in a single observation is ± 3 seconds in Right Ascension and $\pm 3'$ in declination (i.e. ± 1 part in 1300 in the period) in the case of the Cygnus source, as against ± 6 seconds in Right Ascension and $\pm 6'$ in declination in the earlier observations. The accuracy of single observations is similar for all the sources so that few observations were required to reduce the random errors to less than the estimated systematic errors. Between six and twelve observations were made on each.

An overall check of the accuracy of the radio-frequency portion of the equipment may be made very simply. Since the absolute errors in the Right Ascension measurement are inversely proportional to the aerial spacing, a comparison between positions given by the closest and widest spacings gives a good indication of the accuracy of the equipment. Usually no attempt was made to obtain the greatest possible accuracy with the close-spaced aerials, as they are employed only to identify the lobes of the wide-spaced aerials. However, for the accuracy check, six observations were made on the Virgo source, calibrating the two pairs of aerials equally carefully. The difference between

the two positions was 6 seconds, indicating that the likely error in the more accurate position was about $1\frac{1}{2}$ seconds. This is of the same order as the estimated probable error.

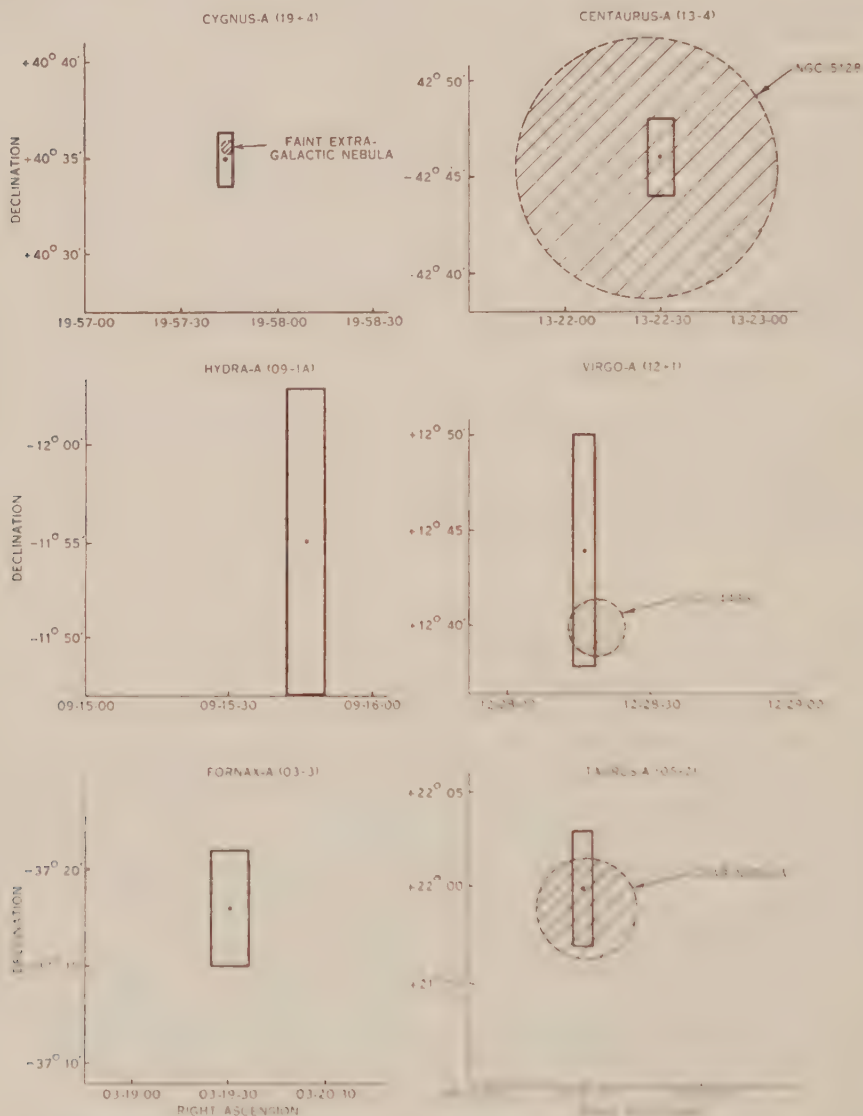


Fig. 1.—Positions of the sources showing the probable error limits, and the approximate outlines of nearby nebulae (Epoch 1950).

IV. THE SOURCES

The sources which have been selected for study are those for which it was expected that an accurate position could be obtained most readily: that is, they are either strong sources ($19+4$, $05+2$, $13-4$, $12+1$) or else weaker sources which show a clear pattern free from pronounced beating effects ($09-1A$, $03-3$).

Their positions are tabulated in Table 1, and in Figure 1 maps of their positions and associated probable errors are shown, together with the positions of nearby nebulae. They are discussed in detail below.

TABLE 1
POSITIONS AND SIZES OF SOURCES

Reference No.	Constellation	Position (Epoch 1950)*		Equivalent Angular Size (minutes of arc)*
		Right Ascension	Declination	
19+4	Cygnus	19 hr. 57 min. 44 sec. $\pm 2\frac{1}{2}$ sec.	$+40^\circ 35' \pm 1\frac{1}{2}'$	<10
13-4	Centaurus	13 22 30 ± 4	$-42^\circ 46' \pm 2'$	25 ± 3
12+1	Virgo	12 28 $15\frac{1}{2} \pm 2\frac{1}{2}$	$+12^\circ 44' \pm 6'$	<10
09-1A	Hydra	09 15 46 ± 4	$-11^\circ 55' \pm 8'$	<15
05+2	Taurus	05 31 29 $\pm 2\frac{1}{2}$	$+22^\circ 00' \pm 3'$	<10
03-3	Fornax	03 19 30 ± 6	$-37^\circ 18' \pm 3'$	20 ± 5

* Errors quoted are the estimated probable errors.

(a) *Cygnus-A* (19+4)

The position of the Cygnus source agrees well with the previous result (Mills and Thomas 1951); also with the most recent position given by the Cambridge workers (Smith 1951). As before, there is no outstanding object in the vicinity with which a likely identification could be suggested. The possibility that the faint nebula noted previously is the source of the radiation is not, however, eliminated by this measurement, particularly as the position of the nebula quoted previously was slightly wrong. The more accurate position is R.A. 19 hr. 57 min. 44.6 sec., Dec. $+40^\circ 35.7'$ (Epoch 1950).*

(b) *Centaurus-A* (13-4)

The identification of this source with NGC 5128 is considerably strengthened by the present observations. Firstly, its position is very close to the centre of the nebula (R.A. 13 hr. 22 min. 26 sec., Dec. $42^\circ 45\frac{1}{2}'$), and secondly, it has a pronounced equivalent angular size which is of the same order as the optical diameter of the nebula.

Because of its finite size, the position of the discrete source refers to an "equivalent" position which, even if the identification is correct, would not necessarily be coincident with the optical centre of the nebula. Also there is an uncertainty introduced into this position by the presence of a nearby source (13-6) which causes a slight modulation of the interference pattern.

Further evidence of the large angular size of the source is given by the fluctuation phenomena. The behaviour appears quite different from that of the

* This position has been quoted by W. Baade (personal communication).

three other strong sources which have been studied, which all show fluctuation effects qualitatively similar to those of the Cygnus source described before (Mills and Thomas 1951). The Centaurus source, however, which has been observed on some 40 occasions, has never shown the intense "fast" fluctuations which are so common with the Cygnus source. Although "fast" fluctuations of about 20 seconds duration have sometimes been observed, they have always been of very low intensity. The most pronounced fluctuations appeared to have periods of between 1 and 2 minutes. If it is assumed that an irregular ionosphere is the cause of the fluctuations the observations could be explained if the source has an angular size which is larger than the angle subtended by the smallest irregularities. Therefore, taking the previous figure of 3 km. as the size of these irregularities and assuming the height is 300 km., the angular diameter of the source would have to be greater than $\frac{1}{2}^{\circ}$. This is consistent with an equivalent angular size of $25'$, as the equivalent angular size is defined on the basis of radiation from a uniformly radiating rectangular strip and can therefore be expected to be less than the actual extent of the source.

(c) *Virgo-A (12+1)*

The identification of this source with the nebula NGC 4486 is also strengthened by the present observations. The nebula position is R.A. 12 hr. 28 min. 18 sec., Dec. $+12^{\circ} 40'$. Because of the low declination the estimated error in the declination measurement is large. The identification cannot therefore be considered to be as reliable as that of the Centaurus source.

(d) *Hydra-A (09-1A)*

Although a relatively weak source, 09-1A appears to be free from trouble due to neighbouring sources as a clear and symmetrical pattern is obtained. The ratio of signal to mean fluctuation is about 10 to 1 when using an 8 second integrating time constant and a bandwidth of 100 kc s., so that an accurate R.A. measurement is possible, although the source is too close to the equator for accuracy in the declination. Actually, the random effects in the measurements are remarkably small, and the main allowances in the estimated errors have been for equipment errors, and for the effect of other sources.

There are no bright nebulae or stars listed near its position, so that there is no obvious identification which can be suggested.

(e) *Taurus-A (05+2)*

The identification of this source with the Crab Nebula (M 1) appears to be strengthened considerably by this measurement, for the position is only $1'$ of arc from the centre of the nebula which has an R.A. of 05 hr. 31 min. 30 sec. and a Dec. of $21^{\circ} 59'$.

(f) *Fornax-A (03-3)*

This is the weakest source investigated. There is some evidence of a modulation of the interference pattern which reduces the accuracy of the measurement slightly, and definite evidence that the source has a comparatively large angular size. Although the source is in a region rich in galaxies, there are no bright nebulae listed at its position.

V. CONCLUSIONS

The idea that the discrete sources are objects of stellar dimensions (radio stars) is not supported by these measurements. All tentative identifications with nebulae made by Bolton, Stanley, and Slee (1949) appear to be confirmed, and in addition the source 03—3 has an apparent angular extension. Thus, of the six sources studied, four appear to be nebulae of some description, leaving the question open in the cases of the Cygnus and Hydra sources.

The question of the galactic or extra-galactic origin of the sources is, also, rather open. The Taurus source appears to be definitely galactic, but it is an unusual radio object (unusual spectrum) and because of its position near the galactic equator it may be expected to belong to the class I sources (Mills 1952) which show a strong galactic concentration. The other nebulae appear to be generally accepted as extra-galactic. Although Evans (1949) has queried this in the case of NGC 5128, some independent observations of Allen and Minkowski (personal communications) confirm the extra-galactic origin. The present measurements are therefore consistent with the suggestion made before that an unusual type of galaxy may be responsible for the randomly distributed class II sources.

VI. ACKNOWLEDGMENTS

The author is indebted to Mr. A. Watkinson and Mr. H. Harant for assistance in the construction and maintenance of equipment and the taking of observations.

VII. REFERENCES

- BOLTON, J. G., STANLEY, G. J., and SLEE, O. B. (1949).—*Nature* **164**: 101.
EVANS, D. S. (1949).—*Mon. Not. R. Astr. Soc.* **109**: 94.
LITTLE, A., and PAYNE-SCOTT, RUBY (1951).—*Aust. J. Sci. Res. A* **4**: 489-507.
MC CREADY, L. L., PAWSEY, J. L., and PAYNE-SCOTT, RUBY (1947).—*Proc. Roy. Soc. A* **190**: 357-75.
MILLS, B. Y. (1952).—*Aust. J. Sci. Res. A* **5**: 266-87.
MILLS, B. Y., and THOMAS, A. B. (1951).—*Aust. J. Sci. Res. A* **4**: 158-71.
SMITH, F. G. (1951).—*Nature* **168**: 555.
STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res. A* **3**: 234-50.

IONOSPHERIC MEASUREMENTS AT OBLIQUE INCIDENCE OVER EASTERN AUSTRALIA

By M. STROHFELDT,* R. W. E. MCNICOL,* and G. DE V. GIPPS*

[Manuscript received February 18, 1952]

Summary

An account is given of night observations of oblique-incidence pulse transmissions on 5.8 Mc/s. over a baseline of 763 km., using a responder technique. The experiment aimed at measuring apparent path lengths as a guide to the identification of the reflecting layer in the ionosphere. The characteristics of beacon triggering are discussed in relation to the echo system received after ionospheric reflection.

A correlation of E_s occurrences as observed at oblique incidence and vertical incidence near the midpoint of the trajectory was effected, and certain characteristics of the E_i layer are presented. Unusual records of Pedersen rays are reproduced and sudden height increases and diffuseness of F_2 echoes are discussed.

A check on the oblique-incidence theory using a Millington transmission curve on vertical-incidence $h'f$ records yielded reasonable agreement between measured and deduced reflection heights, although there is evidence of a tendency for the predicted obliquity factor to be too low. A rough analysis of oblique-incidence penetrations yielded an average value for the frequency separation of the magneto-ionic M.U.F.'s close to half the gyro-magnetic frequency.

I. INTRODUCTION

Recent work (e.g. Appleton and Beynon 1940, 1947; Green 1949, 1950) has indicated that a study of the reflection of radio waves from the ionosphere at oblique incidence may provide information not afforded by the conventional vertical-incidence sounding method. In the present experiments the responder method, as used by Munro (1948) and de Bettencourt and Klemperer (1950), has been applied to transmission paths between Brisbane (lat. 27.5° S., long. 153.0° E.) and Camden (lat. 34.1° S., long. 150.7° E.). These stations are shown on the map in Figure 1. The ground range was 763 km., and the angle of incidence on the E region was thus about 75° and on the F_2 region about 52°. Most of the observations here discussed were made between February 8 and May 13, 1950; but there were some records made before this and some others made in the autumn and winter of 1951. All recording was carried out during the hours of darkness (usually 1830–0530 hours).

By making measurements of the total virtual path with an accuracy of 1.5 km. and knowing the ground range, it has proved possible to measure the virtual height of reflection of the obliquely incident signals with an accuracy of 5 km.

* Physics Department, University of Queensland, Brisbane.

II. DETAILS OF EQUIPMENT

A block diagram of the Brisbane equipment is shown in Figure 2. The Brisbane transmitter emitted pulses at the rate of 50 per sec., each pulse having a power of about 1 kW. and a length of 60 μ sec. The radio frequency used throughout the experiments was 5.80 Mc/s. The output of the receiver at Camden (bandwidth 20 kc/s.) was fed to a trigger unit which was actuated by any pulse whose amplitude exceeded a predetermined minimum. The trigger

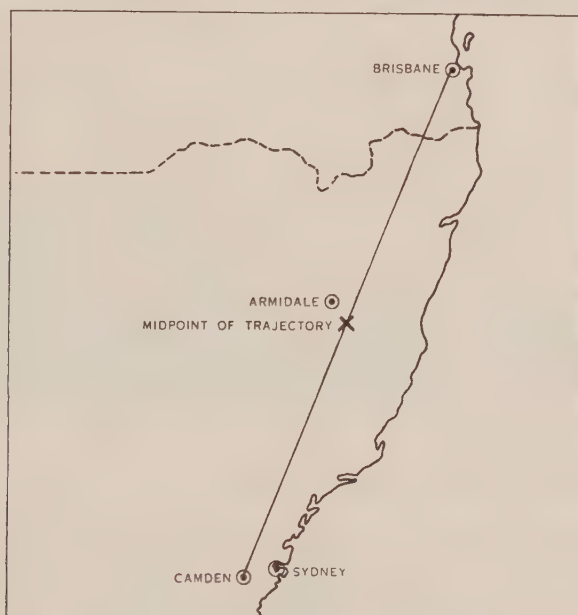


Fig. 1.—Map of eastern Australia showing the locations of the stations.

unit controlled a transmitter (of similar power and pulse length to the one at Brisbane) causing it to emit a pulse. The delay before the emission of the pulse was 400 μ sec. if the level of the incoming pulse was well in excess of the threshold level, but the delay increased to almost 500 μ sec. when the triggering pulse just exceeded the threshold. After the trigger unit operated, it was held off for the next 18 msec. so that, if a string of pulses from Brisbane (travelling via different ionospheric paths) was received each cycle, the Camden equipment would respond only to the first one whose amplitude exceeded the threshold level.

When the noise output from the Camden receiver exceeded the signal output the equipment was triggered at random by noise pulses and successful recording was impossible.

Responses from Camden were received at Brisbane and displayed on a brightness-modulated cathode-ray oscillograph with a linear 50 c/s. time base, whose start was delayed so that only the part representing total path lengths between 1500 and 2500 km. appeared on the screen.

The minimum possible total path, allowing for delay at Camden, slightly exceeded 1600 km. Accurate 50 km. range marks were also supplied for 10 sec. every minute. The pattern was photographed on 35 mm. film moving uniformly at 8 mm./min.

The Camden transmitter was switched off for 5 seconds each minute, thus providing for easy identification of the Camden pulses in the presence of multiple echoes from the Brisbane transmitter. For much of the time the working frequency was above the critical frequency at Brisbane and there were no normal-incidence echoes.

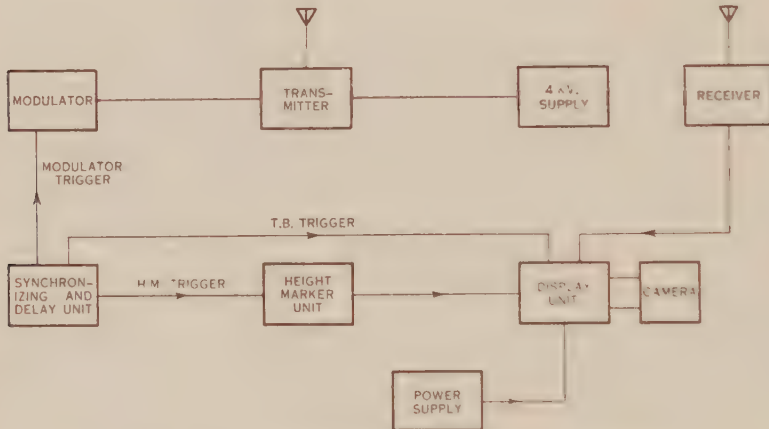


Fig. 2.—Block diagram of the Brisbane equipment.

III. EXPERIMENTAL RESULTS

(a) Nature of Transmission Paths

It is clear that the transmission of radio signals between the two stations can take place via a number of paths, using single and multiple reflections from the F_2 layer, and the E_s layer if present. Combinations of paths involving both these layers are also possible. If pulses due to a number of different propagation modes arrive at Camden at reasonable strength, only the pulse of shortest path will trigger the Camden transmitter, as explained in Section II, and hence the records made at Brisbane will show only one response for each possible mode corresponding to the various return paths only.

Identification of the transmission path concerned for any particular response was based on the assumption that the virtual height of reflection for E_s echoes would be close to 100 km., and for F_2 echoes would be in the range 250–350 km., except during disturbed periods. Sample records are shown in Plate 1, Figures 1 and 2.

Plate 1, Figure 1, illustrates E_s triggering. In it responses with apparent total paths of 1700, 1790, 1860, 2130, and 2320 km. are visible. After subtracting 120 km. to allow for the delay at Camden we are left with the following actual paths:

$E_s : E_s$ (1580 km.); $E_s : 2E_s$ (1670 km.); $E_s : F_2$ (1740 km.); $E_s : 2F_2 - E_s$ (2010 km.); $E_s : 2F_2$ (2200 km.). Note that magneto-ionic splitting shows on this last echo. The diffuse trace between 1980 and 2080 km. does not show the Camden identification breaks. It is a normal-incidence echo from the F_2 region over Brisbane, triply reflected at a virtual height of 330 km.

Plate 1, Figure 2, illustrates F_2 triggering. It shows responses with apparent total paths of 1890 and 2390 km., which are due to $F_2 : F_2$ (1770 km.) and $F_2 : 2F_2$ (2270 km.) actual paths. The trace at 1820 km. is again a triply-reflected normal-incidence F_2 echo.

Any ambiguity in interpretation can be resolved by taking all echoes recorded into consideration. Thus if the *first* response recorded corresponds in path

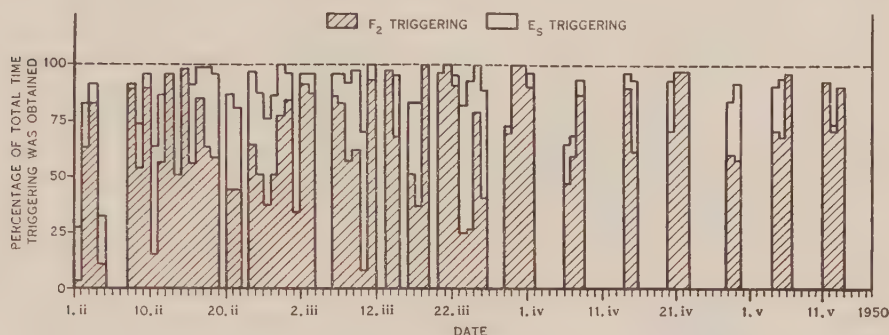


Fig. 3.—Occurrence histogram for the commonest triggering modes (E_s and F_2).

length to either $E_s : F_2$ or $F_2 : E_s$ one would interpret it in the latter way since, if the E_s path gives a sufficient intensity of outgoing signal to cause triggering, it should also give a strong return signal and an $E_s : E_s$ response should be shown on the record, preceding the $E_s : F_2$.

If the E_s signal is fading deeply it will at times be unable to trigger Camden and there will be a repeated switching from E_s to F_2 triggering, as is shown in Plate 2. The changes in delay of the trigger unit, as the strength of the E_s echo drops below the trigger threshold and then rises again, show clearly on this record. When the E_s signal is fluctuating rapidly just at the critical level for triggering a flickering from one path to the other appears on the record, the changes occurring sometimes within a few fiftieths of a second of each other.

By finding the proportion of the records which indicates triggering of any given type, it is possible to compute the proportion of the time for which the strength of the signal reaching Camden via the various paths exceeded about $50 \mu\text{V./m.}$ For the two commonest modes (single F_2 and single E_s) the results (for 1950 only) are shown in the histogram in Figure 3. The shaded portions represent single F_2 , and the clear portions single E_s triggering. As would be expected from the normal-incidence observations of E_s at Brisbane (McNicol and Gipps 1951) the proportion of E_s triggering is greater in summer than later in the year.

Examination of the mean diurnal variation over the whole period suggests that a slight maximum in the amount of E_s triggering occurs at midnight.

This is curious in that a statistical analysis of normal-incidence records of occurrence of E_s at Brisbane shows no evidence of a midnight maximum. Since the total amount of triggering throughout the night was fairly uniform, the increase in E_s triggering around midnight automatically resulted in a decrease in the amount of F_2 triggering at this time, since, if the E_s signal triggered Camden, it was not possible for the F_2 signal to do so as well. A strong F_2 response is, however, almost always recorded amongst the returning pulses under these circumstances. Thus the decrease in amount of F_2 triggering around midnight

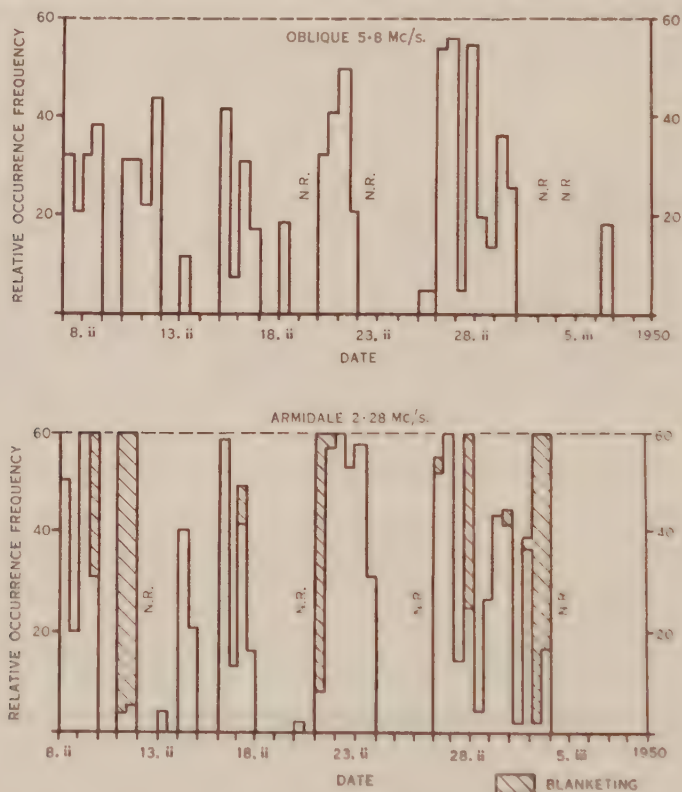


Fig. 4.—Histogram showing the correlation between E_s triggering and E_s occurrence (2.28 Mc/s.) at Armidale. N.R., no record.

does not imply that the F_2 signal was weaker at this time. In fact it is notable that on many nights in autumn, when there was no E_s present, the records show very good F_2 triggering, the responses extending continuously right along the record. An example of such triggering may be seen in Plate 1, Figure 2.

A fairly direct check on the E_s reflections was made by operating a normal-incidence recorder at Armidale, very nearly at the midpoint on the Brisbane-Camden path. For reflections at a virtual height of 100 km., over a ground range of 760 km., the equivalent vertical frequency corresponding to the oblique frequency of 5.80 Mc/s. is 1.63 Mc/s. The normal-incidence recording was made on a frequency slightly greater than this, namely, 2.28 Mc/s., and therefore it is to be expected that there will be a few occasions on which E_s reflections were

recorded at oblique incidence but not at normal incidence. In general, however, the correlation was very good, as shown in Figure 4.

A particular search was made of the records to see if any cases occurred of propagation via reflections from E_s clouds not on the direct line between Brisbane and Camden. Such transmission paths would have shown up as paths slightly in excess of the normal E_s path length but none were found.

Apart from the common E_s and F_2 transmission paths, a few unusual ones were recorded, such as $2E_s$, $E_s + F_2$, and $2F_2 - E_s$ (i.e. an M-type reflection). These, especially the first, have been observed to constitute the triggering path for brief intervals when the main E_s path failed. Such behaviour indicates large spatial variations in reflectivity of the E_s layer.

On the other hand, there is no reliable evidence of $2F_2$ triggering, indicating that this layer is comparatively uniform. (The smaller angle of incidence in the $2F_2$ case would mean a lower probability of reflection, near the M.U.F.)

(b) Behaviour at Penetrations of F_2 Layer

On many records penetrations of the F_2 layer could be observed and in some cases there was a subsequent reappearance of the signal as the F_2 layer ionization density increased again. On some records a succession of penetrations and reappearances is observable, indicating a cyclic fall and rise in ionization density. A typical record is reproduced in Plate 3, Figure 1. Such behaviour suggests the passage over the reflecting point of a series of travelling disturbances with consequent fluctuations in ionization above and below the critical value required to reflect the oblique-incidence signals.

When the working frequency is well below the M.U.F. no separation between the two magneto-ionic components is usually observed. However, as the M.U.F. for the ordinary ray (*o*-ray) is approached, the virtual path for this ray increases resulting in a clear splitting of the received echo. Eventually penetration may occur, subsequent triggering being by the extraordinary ray (*e*-ray). The *e*-ray increases for a time to be reflected quite strongly, and it is obvious from the records that perfectly satisfactory transmission is quite possible by the *e*-ray alone.

The high-angle Pedersen ray was also recorded very well for both magneto-ionic components when the working frequency was near the M.U.F. for the component concerned. During 1951 conditions suitable for recording of the Pedersen ray frequently persisted for long periods (up to 1 hour). An example is shown in Plate 3, Figure 2. When the low-angle ray and the Pedersen ray coalesce, there is no suggestion of beating because of mutual interference, such as has been observed by Appleton and Beynon (1947) when working with continuous waves.

The penetrations and returns are in general quite sharp, with little suggestion of transmission by means of scattering once the working frequency exceeds the M.U.F. It is possible to decide to within, say, 0.1 min. the instant at which the M.U.F. fell below or rose above 5.8 Mc/s. On the few records in which scattering is observable, the vertical-incidence records taken at the same time indicate pronounced diffuseness (see Section III (c)).

(c) Diffuseness of Echoes

Normal-incidence echoes received from the F_2 region at night are often very diffuse, particularly in the case of multiple reflections, for example, see the normal-incidence echoes showing at 1980 km. on Plate 1, Figure 1.

It seems established by Booker and Gordon (1950) that this diffuseness arises from a non-uniform structure of the F_2 layer, probably due in large measure to turbulence. On the other hand, over 95 per cent. of the oblique-incidence records are quite free from diffuseness; this is the case even on records which show marked diffuseness on the simultaneous normal-incidence traces. It is suggested (Gipps and McNicol, unpublished data), from oblique-incidence observations with shorter ground ranges, that diffuseness decreases as the angle of incidence of the wave on the ionosphere increases. The fact that in Plate 1, Figure 1, the doubly-reflected F_2 response shows signs of diffuseness while the singly-reflected F_2 response shows none, is not in itself corroboration of this, however, since, even at (apparent) normal incidence, diffuseness is usually worse on the second reflection than on the first. Dieminger (1951) has accounted for this in terms of oblique reflections from irregularities on the ground.

(d) Sudden Height Changes

When the transmission path is via F_2 a change in height of the reflecting layer produces a nearly equal change in the range of the oblique-incidence response. In this way it has been possible to detect rapid rises in height of the F_2 layer (of the order of 50 km. in 10 to 30 min.) at the reflecting point. On some records normal-incidence echoes also appear and indicate simultaneous changes in height of approximately the same amount for the F_2 layer at Brisbane. Similar rises are also shown on the routine $h'f$ records. (An $h'f$ record takes the form of a graph of the variation of critical frequency, abscissae, with equivalent height of reflection, ordinates, the former being on a logarithmic scale, the latter on a linear scale.)

IV. EXPERIMENTAL CHECKS OF THEORY OF OBLIQUE-INCIDENCE TRANSMISSION

(a) Martyn's Theorem

According to Martyn's theorem the virtual height of reflection at oblique incidence (h'_{ob}) at a given frequency f should be the same as the virtual height of reflection at normal incidence of a frequency $f \sec \varphi$, where φ is the angle of incidence of the wave on the ionosphere. The latter frequency will be called here the "equivalent vertical frequency" (E.V.F.). From the records we can find h'_{ob} , and knowing $\sec \varphi$ we can calculate the E.V.F. Since there was no normal-incidence $h'f$ recorder operating at the midpoint, the value of h' for normal incidence at the E.V.F. had to be deduced from records made at Brisbane and Canberra. This was subject to serious uncertainty, and an accurate check on Martyn's theorem is therefore not possible.

However, several records were analysed using the method developed by Millington (1938). A transmission curve was constructed by plotting on a semi-log transparent graticule the calculated value of h' against $\cos \varphi$ for a

IONOSPHERIC MEASUREMENTS AT OBLIQUE INCIDENCE

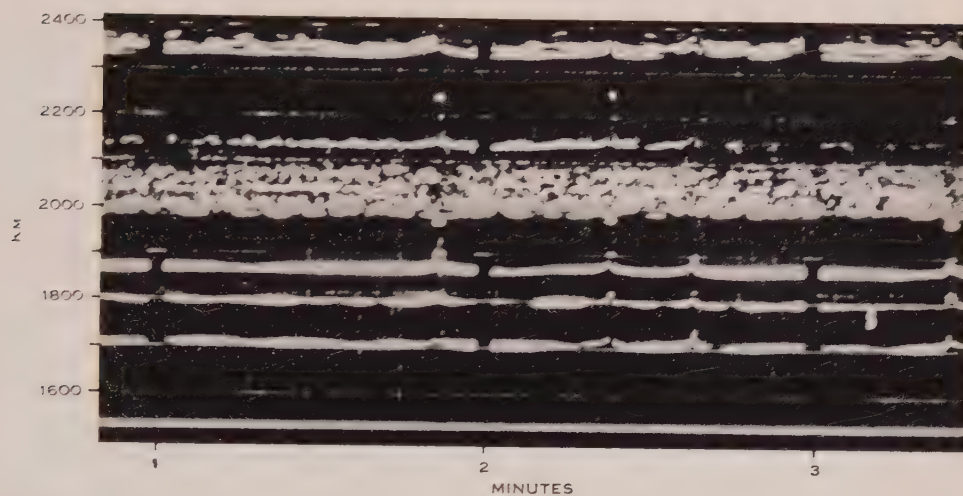


Fig. 1. The E_s triggering mode showing responses here at 1700, 1790, 1860, 2130, and 2320 km. total path lengths. The diffuse trace at 2000 km. is a normal-incidence echo, triply reflected on the F_2 layer.

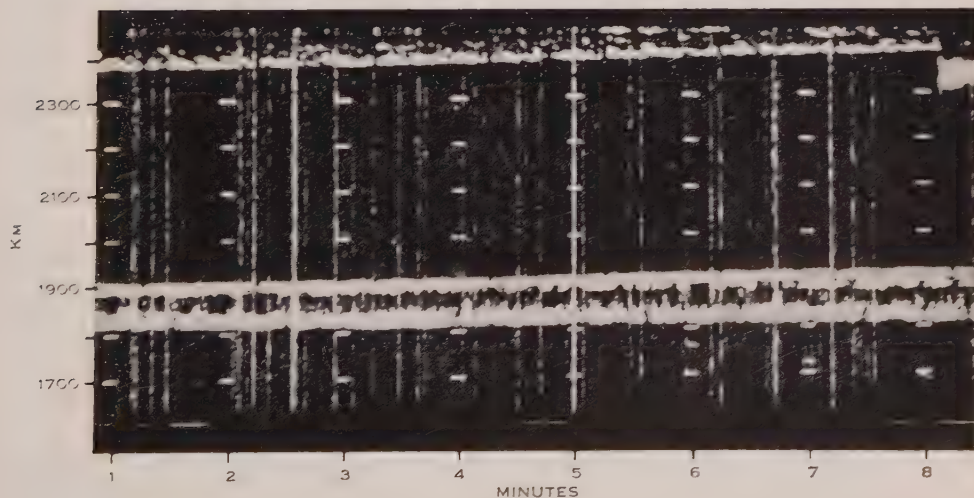
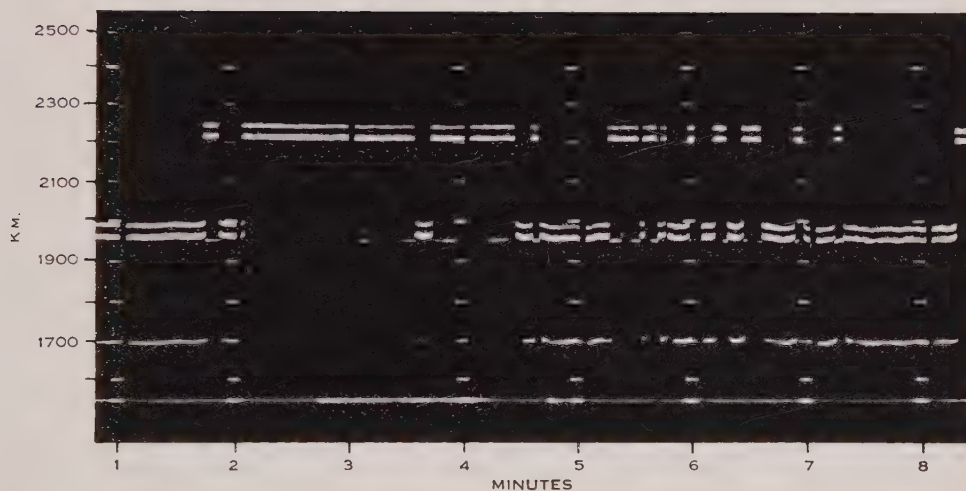


Fig. 2.—A sample record of the F_2 triggering mode with responses at 1890 and 2390 km. total path lengths. The trace at 1820 km. is a normal-incidence echo.

IONOSPHERIC MEASUREMENTS AT OBLIQUE INCIDENCE



Repeated switching between the two chief triggering modes (E_s and F_2) often occurs as in this record. The $E_s : E_s$ and $E_s : F_2$ (o - and e -rays) responses are recorded at 1700, 1900, and 1950 km. respectively corresponding to the E_s mode; while at 2200 and 2240 km. are the $F_2 : F_2$ responses for the F_2 mode, with a weak $F_2 : E_s$ response at 1950 km. Triggering for the F_2 mode is by the extraordinary ray.

IONOSPHERIC MEASUREMENTS AT OBLIQUE INCIDENCE

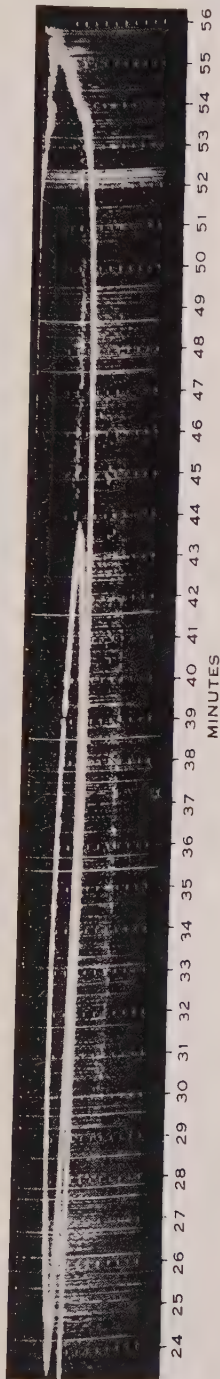
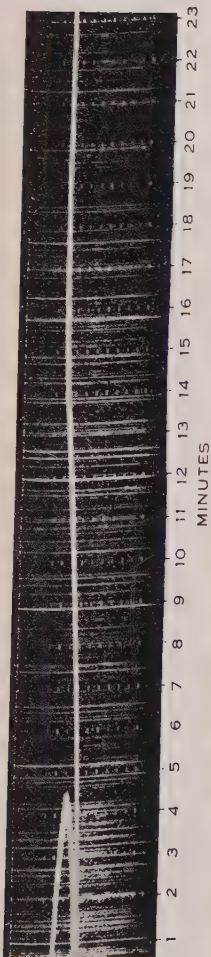


Fig. 1.—A slow pulsation of M.U.F. results in a reappearance of the *o*-ray followed within 20 min. by another penetration of that ray. The *e*-ray corresponding to the lower trace also penetrates 52 min. after the first *o*-ray penetration.

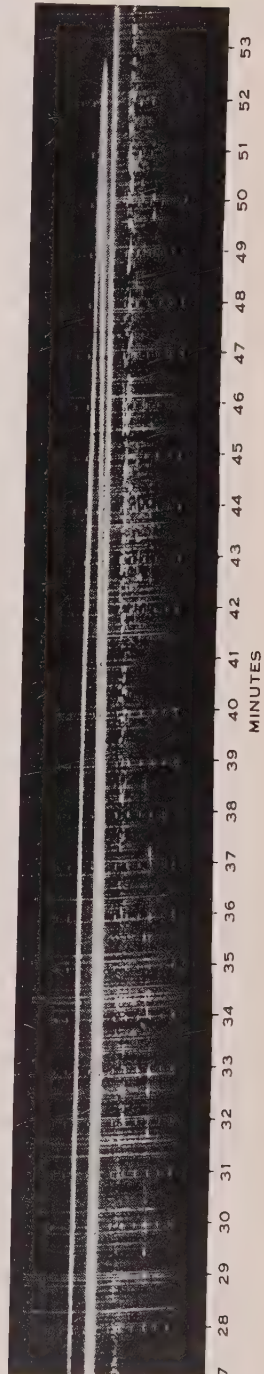
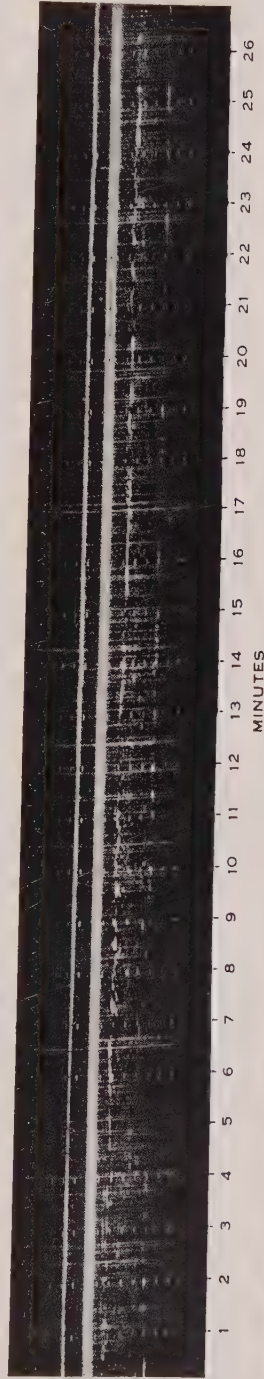


Fig. 2.—An extremely slow change in M.U.F. led to this record of an *o*-ray penetration. The Pedersen ray eventually coalesces with the *o*-ray with no suggestion of interference beating.

great circle distance of 763 km. This calculation is based on a uniform F_2 layer, in which the ionization varies parabolically with height. Allowance is made for the curvature of the Earth. The normal-incidence $h'f$ curve is plotted on a corresponding semi-log graticule and the transparency is slid over this until the $\cos \varphi = 1$ ordinate lies over the working frequency on the $h'f$ scale. The points of intersection of the two curves then give the virtual heights of reflection at normal incidence of the E.V.F.'s. If the working frequency is below the M.U.F. there will be in general two points of intersection corresponding to the low-angle ray and the Pedersen ray. When the working frequency equals the M.U.F., these two rays coalesce and the curves are tangential.

Tested in this way reasonable agreement was found between the measured and deduced heights, but only 12 cases have been examined in detail. Of these, four gave agreement to within 10 km., while the remainder were divided equally between values too high and too low. There was a slight tendency for larger errors to occur in the cases when the actual height of reflection at oblique incidence was higher than the calculated height of reflection at normal incidence than in the reverse cases. This could be viewed as a tendency for the predicted obliquity factor to be too low.

(b) Check on Sellmeyer Formula for Maximum Usable Frequency

Using the transmission curve described in the preceding section, the times when the M.U.F. as deduced from Brisbane $h'f$ records fell below 5.8 Mc/s. were compared with the observed times. Normal-incidence $h'f$ records were made only once every 10 min. Eight oblique-incidence penetrations of the o -ray occurred within 10 min. of the time deduced in this way, and most of the remaining six occurred within 20 min. Once again, no great accuracy can be expected in a check of this nature, since vertical-incidence $h'f$ records, made using the same frequency at three stations spaced only 100 km. apart (Gipps and McNicol, unpublished data), often indicated times of penetrations of the layer differing by several minutes and occasionally up to 1 hour.

(c) Values of Frequency Separation of o - and e -Ray M.U.F.'s

An attempt was made to estimate from the records the difference between the penetration frequencies of o - and e -rays at oblique incidence. By plotting the normal-incidence critical frequencies of the F_2 layer (as measured at Brisbane and Canberra) against time, the rate of change of critical frequency is deduced; then, assuming an obliquity factor (ratio of $f_o F_2$ to (M.U.F.) F_2), it is, in principle, possible to apply this calculated rate of frequency change to a limited period immediately succeeding the o -ray penetration where often an e -ray penetration occurs.

Many difficulties arise in the attempt to apply this method. The obliquity factors are uncertain even when corrected to conform to the observed reflection heights. Also, the rate of change of $f_o F_2$ was in many cases either not constant or different from Brisbane and Canberra.

The average estimated value of the frequency separation of the o -ray and e -ray M.U.F.'s is 0.65 Mc/s., with an uncertainty of at least 0.02 Mc/s. This

agrees with half the gyromagnetic frequency for a true height of reflection of 300 km. (1.32 Mc/s.). It was possible also to make rough estimates based on those cases when only the α -ray penetration was observed. These gave consistent results.

V. ACKNOWLEDGMENTS

This work has been carried out at the University of Queensland as part of the programme of the Radio Research Board, C.S.I.R.O. The authors are indebted to Professor H. C. Webster, Physics Department, University of Queensland, for much helpful discussion and criticism, to Dr. G. H. Munro, Radio Research Board, and to Mr. J. A. Somerville, New England University College, for cooperation in the experimental programme.

VI. REFERENCES

- APPLETON, E. V., and BEYNON, W. J. G. (1940).—*Proc. Phys. Soc.* **52**: 518.
APPLETON, E. V., and BEYNON, W. J. G. (1947).—*Proc. Phys. Soc.* **59**: 58.
DE BETTENCOURT, J. T., and KLEMPERER, H. (1950).—*Proc. Inst. Radio Engrs. N.Y.* **38**: 791.
BOOKER, H. G., and GORDON, E. W. (1950).—*Proc. Inst. Radio Engrs. N.Y.* **38**: 401.
DIEMINGER, W. (1951).—*Proc. Phys. Soc.* **64B**: 142.
GREEN, A. L. (1949).—"Supplementary Handbook for use with the Radio Propagation Bulletin."
(Ionospheric Prediction Service of the Commonwealth Observatory, Australia, June and October.)
GREEN, A. L. (1950).—"Supplementary Handbook for use with the Radio Propagation Bulletin."
(Ionospheric Prediction Service of the Commonwealth Observatory, Australia, June.)
McNICOL, R. W. E., and GIPPS, G. de V. (1951).—*J. Geophys. Res.* **56**: 17.
MILLINGTON, G. (1938).—*Proc. Phys. Soc.* **50**: 801.
MUNRO, G. H. (1948).—*Nature* **162**: 886.

THE GROWTH OF CLOUD DROPS BY CONDENSATION

II. THE FORMATION OF LARGE CLOUD DROPS

By P. SQUIRES*

[*Manuscript received August 3, 1951*]

Summary

The observed stability of many deep non-freezing clouds suggests that there is a gap between the sizes normally reached by droplets due to condensation and the sizes at which coalescence of drops becomes effective in producing rain. One way such clouds could yield rain would be for condensation to give rise to some drops bigger than usual.

Conditions for the growth of some drops to a relatively large size by condensation are investigated theoretically. This requires that in some air parcels an undue proportion of the available water vapour should condense on relatively few nuclei. The types of vertical motion and nucleus spectra which could yield this result are investigated.

Slow motions are favourable. Suitable spectra are: (i) a few "giant" nuclei ($M = 10^{-10}, 10^{-12}$); (ii) all but a few nuclei relatively small. Rain itself tends to produce type (ii) spectrum, so that it could perhaps act regeneratively.

It is found that very large cloud drops can be formed by condensation under conditions which, on present knowledge, seem likely to occur at times, but not normally. The hypothesis that the factors studied here largely determine colloidal stability in non-freezing cloud is therefore satisfactory, inasmuch as it leaves room for both stability and instability.

I. INTRODUCTION

(a) Observations of Rain from Non-freezing Cloud

It is now well established that rain sometimes falls from clouds which are wholly warmer than 0°C .; observations bearing on this point have been reported by various workers (Heywood 1940; Wexler 1945; Kotsch 1947; Craddock 1949; Hunt 1949; Davies 1950; Kinshott 1950; Virgo 1950; Smith 1951).

This obviously cannot be explained by the generally accepted theory of Bergeron, according to which most precipitation originates from ice crystals. Clearly, it is important for the understanding of cloud and precipitation to account for such cases, not only because of their intrinsic interest, but also because the same processes may operate even when precipitation falls from a cloud in part colder than freezing.

In the cases described by the authors cited, there appears to have been nothing obviously unusual about the clouds which yielded the rain. Many clouds remain colloiddally stable which are similar to those described by the authors mentioned in respect of temperature, depth, and turbulence.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

Now, it is possible that the release of colloidal instability depends on some delicate concatenation of circumstances. But it is natural to try first to see whether there is not some critical factor or factors which, when non-freezing clouds yield rain, have characters quite different from those which commonly occur in temperate latitudes, where such clouds normally remain stable. The only common feature among the observed instances of rain from non-freezing cloud is that they all occurred either over or near the sea: most of them in tropical latitudes. There is a very general impression among tropical meteorologists that this type of rain is quite common over and near tropical seas. This is confirmed by Leopold and Halstead (1948), Davies (1950), Foster (1950), and Virgo (1950). Kraus (1949), Woodcock and Gifford (1949), and Woodcock (1950) have suggested that such rain may be due to the presence of very large sea-salt nuclei. Certainly, the spectrum of nuclei would be different over land and sea, and may perhaps vary with latitude.

It seems possible, therefore, that variations in the spectrum of condensation nuclei may be one of the factors which determine the stability of a water cloud. This does not of course exclude other factors: a spectrum which is suitable for the production of colloidal instability may need to be combined, for instance, with a particular sort of motion.

(b) *The Mechanisms of Drop Growth*

If we now look at this problem from the theoretical angle, it is obvious that, in a cloud devoid of ice crystals, raindrops must form by the action of condensation or coalescence, or both. Evidently the former process operates first to produce the initial cloud droplets. It will be shown that, on any reasonable assumptions, condensation probably cannot contribute to significant growth much beyond the conventional upper limit of cloud drops, $r = 100\mu$. If this is so, the drops in any precipitation heavier than drizzle which falls from non-freezing cloud must be formed by the action of coalescence. Wright (1940) has mentioned that the average chlorine concentration is much the same in cloud and rain: this suggests that raindrops consist mostly of swept-up cloud drops, even in the case of Bergeron type rain. On the other hand, at drop sizes below $r = 100\mu$, little is known with certainty about coalescence, except that its efficiency increases with the size of the drops. The calculations made by Langmuir (1948) on collection efficiencies, and the reasoning used by Swinbank (1947) concerning the possibility of "bounce off" at small sizes both point this way. Even with constant collection efficiency, the mass rate of growth of a cloud drop varies as r^4 approximately. Thus coalescence must proceed more effectively when large drops are already present.

It has already been suggested that the condensation nucleus spectrum might be important in this connection, but it can have an influence only during the condensation phase. These two lines of evidence clearly suggest that it is desirable to investigate the conditions which favour the formation of large cloud drops by condensation, and to form an estimate of the maximum size to which these drops can grow. A theoretical analysis of these questions is presented in this paper.

II. CONDENSATION OCCURRING CHIEFLY ON A RESTRICTED NUMBER OF NUCLEI

(a) *General*

Several calculations have been made of the sizes which can be reached by condensing drops. These have for the most part been based on assumed supersaturations and growth times. As was pointed out in Section IV (a) of Part I of this series (Squires 1952), there is no valid basis in observation for assuming an upper limit to the supersaturation. As will be seen below, the crux of our problem lies rather in the fact that only a limited mass of water is available. It is the distribution of this water among the nuclei present which is significant. Under the adiabatic assumption, which will be made here, this depends on the nucleus spectrum and the vertical motion. No discussion which omits these factors can be truly relevant to this problem.

The most obvious situation in which large drops would form is one where the total number of condensation nuclei is small. Normally the concentration of nuclei as measured by an Aitken counter is of the order of thousands per cc. Byers (1949), however, has pointed out that some very small nucleus-counts have been obtained over the open sea, and suggests this as an explanation of rain from non-freezing clouds. There are not enough data to determine whether the frequency of sufficiently low nuclear concentrations is such as to explain the apparently rather frequent occurrence of such rain over and near tropical seas. The evidence of Woodcock (1950) is not very favourable to this view. It would be natural to suppose that a cloud consisting of only a few drops per cc. would look rather different from ordinary clouds. It might, however, be speculated that the tropical phenomenon "serein" (rain from a clear sky), mentioned, for example, by Wood Jones (1934) could arise in this manner.

On the other hand, if large drops are to be formed by condensation in the presence of some hundreds or thousands of nuclei per cc., some principle of selection must operate, for only a few tens of large cloud drops can be formed per cc. in any reasonable lifting. It is evident that this must be related to the nucleus sizes, for the largest nuclei present must always form the biggest drops.

The heterogeneity of the nucleus spectrum is therefore of the essence of the problem. It is natural, therefore, to seek some simplification by taking the vertical velocity as uniform. The work of Howell (1949) and Kraus and Smith (1949) has, however, shown that, with such a motion, even a very heterogeneous group of nuclei gives rise to a markedly homogeneous cloud. Moreover, as we are concerned rather to find what spectra of nuclei and what motions produce a certain result, namely, the concentration of most of the liquid water on a few of the larger nuclei, it is very undesirable to particularize either the spectrum or the motion more than is necessary. Furthermore, our knowledge of motions in cloud, and, more especially, of nucleus spectra, is so slight that special assumptions about these factors should be avoided as far as possible, for none that we can make at present would have any real plausibility. It seems best at this stage to make assumptions only about the form of the spectrum

and the motion, so that they can be defined by a reasonably small number of parameters, and then find what values of these parameters will lead to the required result.

An indication of a fertile approach is given by Figure 5 of Part I. These curves refer to growth at constant supersaturation. They show that if there are no "giant" nuclei ($M=10^{-12}$ to 10^{-10}) present, and if large cloud drops are to be formed in a period of less than 10^4 seconds, the supersaturation must have a value at which the threshold effect enters; that is, even a continuous spectrum of nuclei will tend to behave as if it consisted of only two size-groups, the smaller nuclei remaining unactivated and the larger ones forming the large drops. This suggests that the influence of heterogeneity in the nucleus spectrum could be investigated by considering only two groups of nuclei and combining this with a vertical motion so constructed as to ensure that the group of smaller nuclei remains unactivated. Obviously, the number of larger nuclei must not be too great; otherwise the liquid water formed in any reasonable lifting will be inadequate to form large drops on them. Of course the supersaturation, S , is not constant in actual cloud, but it must be expected that the general trend will be for it to decrease with height, once the activation stage is past. This was found by Howell (1949) and Kraus and Smith (1949) in calculations made with constant velocity. It follows in a more general way from equation (4.3) of Part I, for Z_1 will increase as the drops grow, so that inordinate velocities would be required in the body of the cloud to cause S to increase above the highest value which it reached in the activation stage. Thus nuclei which are not activated early in the vertical motion are likely to remain inactive throughout and a continuous spectrum will tend to behave like one containing only two sizes of nuclei even though S varies. This tendency was found in the numerical examples worked by Howell (1949), although the final drop sizes were not particularly big.

If "giant" nuclei ($M=10^{-12}$ to 10^{-10}) are present, the suggestion drawn from Figure 5 of Part I no longer holds. Nevertheless, it will be seen that some account can be taken of such nuclei. It will appear later that there is reason to believe that, at any rate in cases of widespread and long continued rain, such nuclei are probably not important, for it seems likely that they are very effectively removed by rain from the air entering the cloud from below.

As regards the vertical motion, it will be seen that it is necessary only to satisfy an inequality which limits the upward velocity at each level. This allows much latitude in choosing the form of the motion. Further, it must be noted that we are dealing with the motion of an individual air parcel only. Nothing is assumed about the motion in the rest of the cloud, which may be entirely different from that in the parcel considered. It will be sufficient if, now and then, somewhere in the cloud a parcel moves upwards in such a way as to satisfy the inequality, through a layer deep enough to yield an adequate release of liquid water. For these reasons, the condition on the motion of the air parcel cannot be regarded as difficult to meet, in general. The large drops produced in this particular parcel may be supposed to mix with the surrounding

cloud at an upper level and may perhaps grow further by coalescence as they fall through the cloud. The theory therefore does not necessarily imply that a pure large-drop spectrum could be observed in the upper part of the cloud.

We will therefore consider first the movement of a parcel of air in which only two groups of nuclei are present. The result will be to define a *sufficient* set of conditions for the formation of drops of a given size. In the process of deriving these quantitative conditions, various assumptions are made in order to keep the treatment reasonably simple. These are not such as to render the treatment unreal, and may be accepted as plausible in a calculation which is more a generalized illustration than a demonstration.

As in Part I, all numerical values of physical quantities are taken at 10 °C., 800 mb., unless otherwise stated. Further, as we are concerned with a phenomenon occurring at cloud levels, and as far as the observations go, in maritime air, it will be assumed that the total number of nuclei per gram of dry air does not exceed 10^8 ($\approx 10^5$ per cc.). A list of symbols is given at the end of the paper. Symbols common to Parts I and II are not repeated.

(b) Two Categories of Nuclei

Suppose there are n_1 nuclei of size M_1 , n_2 of size M_2 , per gram of dry air, and no others ($M_1 > M_2$). Let S_1 and S_2 be their critical supersaturations ($S_2 > S_1$). Then there will be only two sizes of drop present, whatever the motion. Suppose further that, at the initial level z_0 , $S \leq S_2$, and the M_2 nuclei are in the haze stage, either because the supersaturation has never been high enough to activate them, or as a result of evaporation in a descending current. These nuclei will never be activated, provided S does not exceed S_2 . Let the mass of liquid water associated with the M_1 nuclei be w_1 , with the M_2 nuclei, w_2 . Suppose now that the M_2 nuclei are removed, and a motion is constructed which ensures that S does not exceed S_2 . If with the same motion the M_2 nuclei are re-introduced, S will still remain less than S_2 , for the effect of the M_2 droplets is merely to make the variations of S more sluggish. In particular if S should ever equal S_2 , they will be growing, and hence will tend to reduce S . A motion will therefore be found that keeps S less than S_2 in the absence of the M_2 nuclei. The argument of this Section is not changed in any way if the smaller nuclei are heterogeneous. Then M_2 must be taken as the size of the largest of them, and S_2 as the corresponding smallest critical supersaturation.

As the M_1 droplets all have the same radius, the χ 's in equation (4.2) of Part I can be expressed by simple products, and multiplying by v we find

$$\frac{dS}{dt} = q_1 v - 4\pi E q_2 n_1 F(R_e, \sigma') (Sr - \beta + \gamma M_1 r^{-2}),$$

where r is the radius of the M_1 droplets.

Consider a motion in which $S < S_2$ initially and

$$v < 4\pi E q_2 n_1 (S_2 r - \beta + \gamma M_1 r^{-2}) / q_1, \quad \dots \dots \dots (2.1)$$

at all levels.

By the definition of S_1

$$S_1 r \geq \beta - \gamma M_1 r^{-2},$$

for all (positive) r ; since $S_2 > S_1$, the right-hand side of (2.1) is necessarily positive.

Then since $F(R_e, \sigma') > 1$,

$$\frac{dS}{dt} < 4\pi E q_2 n_1 F(R_e, \sigma') (S_2 - S) r.$$

Hence S can never exceed S_2 . In a motion obeying the condition (2.1), if the M_2 nuclei are now introduced, they will never be activated. Equation (3.1) of Part I shows that the value of r at any level is given by the relations

$$(w_1 + w_2) q_2 + S - q_1 z = W_0 q_2,$$

where W_0 is a constant for any given parcel of air, and

$$w_1 = 4\pi \rho' n_1 r^3 / 3.$$

If the origin of z is chosen at the level where with infinitely slow vertical motion S vanishes (i.e. practically at the cloud base), it is seen that W_0 is a function of the nuclei present. It is the mass of liquid water associated with the M_1 and M_2 nuclei in equilibrium at $S=0$.

From the above equations,

$$r^3 = 3(q_1 z - S + (W_0 - w_2) q_2) / 4\pi \rho' n_1 q_2.$$

Under certain conditions, $W_0 = w_2$; this will be so if the mass of liquid water associated with all the nuclei in equilibrium at $S=0$ exceeds that associated with the M_2 nuclei at their critical radius. Since

$$r_1 = \sqrt{\gamma M_1 / \beta}, \quad r_2 = \sqrt{3\gamma M_2 / \beta},$$

this would follow provided

$$n_1 M_1^{3/2} + n_2 M_2^{3/2} > 3\sqrt{3} n_2 M_2^{3/2},$$

that is

$$M_1 / M_2 > 2.6 (n_2 / n_1)^{2/3}. \quad \dots\dots\dots (2.2)$$

Again, if M_2 is small, w_2 will also be small. It is convenient to exclude cases where z_0 , the minimum value of z , is less than about 150 m. (above the cloud base) and to assume that the maximum value of S_2 is 1°C., which corresponds to $M_2 \approx 10^{-20}$. Since it is postulated that all but a few of the nuclei are of size M_2 or smaller and the total range of nucleus sizes is 10^{-10} to 10^{-22} , this limit on S_2 is not unreasonable. Then $(q_1 z - S)$ can never be less than 0.19°C. since $S < S_2$.

Now, since n_2 is assumed less than 10^8 ,

$$\begin{aligned} q_2(w_2 - W_0) &< 4\pi \rho' n_2 q_2 M_2^{3/2} (\eta_c^3 - \eta_s^3) / 3, \\ &< 1.61 \times 10^{24} M_2^{3/2}. \end{aligned}$$

The proportional error in r caused by neglecting this term therefore cannot exceed $0.28 \times 10^{27} M_2^{3/2}$ per cent. This is less than 1 per cent. if

$$M_2 < 1.1 \times 10^{-18}. \quad \dots\dots\dots (2.3)$$

Provided one of these conditions (2.2) or (2.3) holds, we may write without appreciable error

$$r^3 \geq 3(q_1 z - S)/4\pi\rho' n_1 q_2.$$

As mentioned above,

$$S_1 r \geq \beta - \frac{\gamma M_1}{r^2}.$$

If r assumes values up to the critical radius of an M_1 droplet, this relation will become an equation there. But if M_1 (and hence r_c) is very large, this may not happen. If we consider only growths up to a maximum radius of r_m ,

$S_1 - \frac{\beta}{r} + \frac{\gamma M_1}{r^3}$ will have a positive non-zero lower bound (ψ), provided

$$r_c = \eta_c M_1^{\frac{1}{3}} = 1.060 \times 10^4 M_1^{\frac{1}{3}} > r_m.$$

If $M_1 \leq r_m^2/\eta_c^2$, ψ will be zero. For $M_1 > r_m^2/\eta_c^2$, ψ is a function of M_1 and r_m :

$$\psi = \xi_c M_1^{-\frac{1}{3}} - \beta r_m^{-1} + \gamma M_1 r_m^{-3}. \quad \dots\dots\dots (2.4)$$

The relation (2.1) will then be satisfied provided

$$v < 4\pi E q_2 n_1 (S_2 - S_1 + \psi) r / q_1.$$

If we now define

$$r_f^3 = 3(q_1 z - S_2)/4\pi\rho' n_1 q_2, \quad \dots\dots\dots (2.5)$$

$$r \geq r_f \text{ at all levels,}$$

and if further

$$v_f = 4\pi E q_2 n_1 (S_2 - S_1 + \psi) r_f / q_1, \quad \dots\dots\dots (2.6)$$

the relation (2.1) will be satisfied if

$$v < v_f \quad \dots\dots\dots (2.7)$$

at all levels, the M_2 nuclei will never be activated, and the M_1 droplets will have radii of at least r_f . Equations (2.5) and (2.6) define v_f as a function of z and ψ , the latter being itself a function of M_1 and r_m . The term ψ in (2.6) allows for higher velocities when M_1 is very large. These increased velocities are permissible because, for such large values of M_1 , the vapour pressure over the drops is still appreciably reduced by the effect of the nucleus even when $r = r_m$.

Evidently the ψ term is important for M_1 large and r_m small. For the ranges of M_1 and r_m shown in Table 1, $\psi \leq 0.80$ °C. If (2.5) is always to yield positive values for r_f , it is necessary that

$$q_1 z_0 - S_2 > 0. \quad \dots\dots\dots (2.8)$$

This is assured by the limitations already assumed regarding the possible values of z_0 and S_2 .

In order to express quantitatively a set of conditions suitable for the formation of large drops, it is necessary to choose some type of motion and then express (2.7) in terms of the parameters defining the motion. This inequality limits the vertical velocity at each level, but permits it to increase upwards as the absorptive power of the cloud drops for water vapour, measured by

χ_1 ($\approx n_1 v$), increases. However, a motion with continually increasing velocity is both unrealistic and ill-suited to our problem. Since we are concerned to find how large drops can be produced inside a cloud, it is necessary that the vertical velocity should ultimately decrease to a low value at a level still inside the cloud.

TABLE 1
 ψ AS A FUNCTION OF M_1 AND r_m ($^{\circ}\text{C}.$)

r_m (μ)	M_1				
	10^{-10}	10^{-11}	10^{-12}	10^{-13}	10^{-14} and Smaller
20	0.80	0.080	0.0073	0.0003	0
40	0.10	0.0097	0.0006	0	0
70	0.018	0.0018	0.0001	0	0
100	0.0063	0.0005	0.0000	0	0

In order to reach simple quantitative results, the inequality will be applied to the case of sinusoidal vertical motion, such as might result from oscillations about an equilibrium position, or the presence of vortices with more or less horizontal axes. Such a motion is evidently quite suitable for the present purpose, since the velocity at first increases upwards and then decreases to zero. Evidently it is not necessary to suppose that the motion is accurately sinusoidal, since only an inequality needs to be fulfilled.

III. SINUSOIDAL MOTION

(a) *The Conditions on the Motion with Two Categories of Nuclei*

Consider a motion specified by

$$z - z_0 - A = A \sin 2\pi t/P.$$

The particle will be at its lowest position $z = z_0$ at $t = P/4$, at its mean position $z = z_0 + A$ at $t = 0$, and at its apex $z = z_0 + 2A$ at $t = P/2$. The vertical velocity, v , may be expressed as a function of $\zeta = (z - z_0)/A$

$$v = 2\pi A (2\zeta - \zeta^2)^{1/2} / P.$$

It will be assumed that the initial condition $q_1 z_0 + S_2 > S$ holds at the level z_0 , which will be taken as the initial level. Then v_1 defined by (2.6) above, may be expressed in terms of ζ as follows:

$$v_1 = \left(\frac{4\pi n_1 q_2}{q_1} \right)^{2/3} \left(\frac{3A}{\varphi'} \right)^{1/3} E(S_2 - S_1 + \psi) (\zeta + f_1)^{1/3},$$

where

$$f_1 = (q_1 z_0 - S_2) / q_1 A,$$

which by (2.8) must be positive. Hence v will not exceed v_f at any level provided

$$\frac{A^{2/3}}{P} \leq \left(\frac{6}{\pi\rho'}\right)^{1/3} \left(\frac{n_1 q_2}{q_1}\right)^{2/3} E(S_2 - S_1 + \psi) \frac{(\zeta + f_1)^{1/3}}{(2\zeta - \zeta^2)^{1/2}},$$

for $0 \leq \zeta \leq 2$.

Since f_1 is positive, the function of ζ on the right has a single turning point in this range, where it takes on its minimum value. This value increases with f_1 , as shown in Table 2.

TABLE 2

MINIMA OF $(\zeta + f_1)^{1/3}/(2\zeta - \zeta^2)^{1/2}$ IN THE RANGE $0 \leq \zeta \leq 2$ FOR VARIOUS POSITIVE VALUES OF f_1

f_1	0	0.1	0.2	0.5	1.0	2.0
Minimum	0.916	0.968	1.011	1.114	1.241	1.426

The condition on the motion is most stringent when $f_1 = 0$. Hence on substituting numerical values for 10 °C., 800 mb., a sufficient condition is

$$\begin{aligned} \frac{P}{A^{2/3}} &\geq \left(\frac{\pi\rho'}{6}\right)^{1/3} \left(\frac{q_1}{n_1 q_2}\right)^{2/3} \left| 0.916 E(S_2 - S_1 + \psi), \right. \\ &= 97.54/n_1^{2/3} (S_2 - S_1 + \psi). \end{aligned} \quad (3.1)$$

Provided the initial conditions are fulfilled, any motion obeying this condition will leave the M_2 nuclei unactivated and will produce drops on the M_1 nuclei whose radii at any level will exceed r_f ; in particular at the apex they will have radii of r_a where

$$r_a^3 = 3\{q_1(2A + z_0) + q_2(W_0 - w_2) - S_a\}/4\pi\rho'q_2n_1, \quad (3.2)$$

where S_a is the value of S at the apex. Either (2.2) or (2.3) being assumed, we may write

$$r_a^3 \geq r_f^3 = 3\{q_1(2A + z_0) - S_2\}/4\pi\rho'q_2n_1 > 6Aq_1/4\pi\rho'q_2n_1, \quad (3.3)$$

by (2.8).

The lower bound for r_a given by (3.3) is satisfactory when the level z_0 lies near the cloud base (but not closer than allowed by (2.8)). But when z_0 is comparable with $2A$, it yields only a very rough lower bound.

Near the apex of the motion S becomes small for the upward velocity is small, and S tends to come nearly into equilibrium with the cloud drops. It is shown in Appendix I that, provided $A > 200$ m., the error in r_a due to neglecting S_a in (3.2) cannot exceed 8 per cent., and that this error decreases very quickly with increasing A . As the purpose here is merely to find the conditions that favour the formation of large drops, this error may reasonably be neglected. Leaving aside therefore the more complex cases where $A < 200$ m., we may write without serious error

$$r_a^3 \geq r_b^3 = 3q_1(2A + z_0)/4\pi\rho'q_2n_1. \quad (3.4)$$

In other words, the size of the M_1 drops at the apex may be computed by assuming that they divide equally the water released during lifting from the cloud base. Only a negligible proportion remains in the vapour form as supersaturation when A exceeds a few hundred metres and $S_2 < 1$.

If the air parcel continues to move in a sinusoidal manner and descends from the apex $z_0 + 2A$ to z_0 once more, S will remain less than S_2 , for the argument does not require that r be positive. At z_0 , the M_1 droplets will have partly evaporated and the initial conditions will be re-established. Successive cycles would reproduce, essentially, the events of the first.

(b) Relation between the Size of the Nuclei, the Period, and the Final Drop Size

If a considerable body of air is executing a sinusoidal type of motion in a vortex or in a vertical oscillation, all parts will have much the same period, but the amplitudes will vary widely. Suppose that some parcel of air has its apex far enough above the condensation level to ensure that (3.4) will be satisfied for some given values of r_b and n_1 . Then, by (3.1) and (3.4), the M_2 nuclei will not be activated and the M_1 drops will then have radii at least as large as r_c , provided

$$P > \frac{2^{1/3} \pi \rho'}{3 \times 0.916 E} \frac{r_b^2}{(S_2 - S_1 + \psi) \left(1 + \frac{z_0}{2A}\right)^{2/3}} \\ = \frac{2 \cdot 19 \times 10^7 r_b^2}{(S_2 - S_1 + \psi) \left(1 + \frac{z_0}{2A}\right)^{2/3}} \dots \dots \dots (3.5)$$

that is, if

$$(S_2 - S_1 + \psi) > 2 \cdot 19 \times 10^7 r_b^2 / P \left(1 + \frac{z_0}{2A}\right)^{2/3}.$$

When $z_0 \ll 2A$, this simplifies to

$$(S_2 - S_1 + \psi) > 2 \cdot 19 \times 10^7 r_b^2 / P.$$

The effect of this condition on the possible values of $S_2 - S_1 + \psi$ is shown in Table 3 for the case $z_0 \ll 2A$. When z_0 is comparable with $2A$, these values can be reduced in the ratio $(1 + z_0/2A)^{-2/3} : 1$, for example, for $z_0 = 9A$, about three times.

The values in this table have to be compared with the range of S_2 , S_1 , and ψ , namely, 10^{-5} to 1 °C. for S_1 and S_2 , and 0.0-8 °C. for ψ . It is clear that the production of very large drops such as $r = 100\mu$ would be assured only with long periods (P large) and values of S_2 close to the upper limit. (At $r_b = 100\mu$, ψ (Table 1) is negligible compared with these limits, even for "giant" nuclei.) It is evident that these conditions could occur only rarely, for the majority of the nuclei in the air have to be of size M_2 (or smaller) with critical supersaturations S_2 or larger. Moderately large cloud drops would be formed under much less stringent conditions. At $r_b = 40\mu$, ψ is still small compared with

these limits, but the required value of S_2 is not so large, so that it is less improbable that the majority of the nuclei could have critical supersaturations of this size or larger.

TABLE 3

LOWER LIMITS ON THE AVAILABLE SUPERSATURATION ($S_2 - S_1 + \psi$) IN °C. WHICH ENSURE THE PRODUCTION OF DROPS OF RADIUS r_b AT THE APEX OF A SINUSOIDAL MOTION OF PERIOD P
Height of initial level above cloud base small compared with the amplitude

r_b (μ)	P (sec.)				
	500	1000	2000	5000	10,000
20	0.18	0.09	0.04	0.018	0.009
40	0.70	0.35	0.18	0.070	0.035
70	2.15	1.07	0.54	0.21	0.11
100	4.38	2.19	1.10	0.44	0.22

For the larger values of r_b , ψ is always small compared with the limits in Table 3. If the condition of Table 3 is to be met, S_1 will usually be small, for the requirement on S_2 is already quite stringent. In that case, Table 3 may be taken as giving a lower limit to S_2 . Again, if $\psi > S_1$ (r_b not too great, and M_1 large) it is sufficient if S_2 exceeds the lower limits given in the table. As S_2 is a function of M_2 , this gives an upper limit to M_2 . The values of this limit, for $z_0 \ll 2A$, are given in Table 4. When z_0 is comparable with $2A$, these values can be increased in the ratio $(1 + z_0/2A)^{4/3} : 1$, for example, for $z_0 = 9A$, about 10 times.

TABLE 4

UPPER LIMITS ON THE SIZE OF THE SMALLER NUCLEI (M_2) WHICH ENSURE THE PRODUCTION OF DROPS OF RADIUS r_b AT THE APEX OF SINUSOIDAL MOTIONS OF PERIOD P
Height of initial level above cloud base small compared with the amplitude

r_b (μ)	P (sec.)				
	500	1000	2000	5000	10,000
20	3.8×10^{-19}	1.5×10^{-18}	6.1×10^{-18}	3.8×10^{-17}	1.5×10^{-16}
40	2.4×10^{-20}	9.5×10^{-20}	3.8×10^{-19}	2.4×10^{-18}	9.5×10^{-18}
70	2.5×10^{-21}	1.0×10^{-20}	4.1×10^{-20}	2.5×10^{-19}	1.0×10^{-18}
100	6.1×10^{-22}	2.4×10^{-21}	9.7×10^{-21}	6.1×10^{-20}	2.4×10^{-19}

The general conclusions drawn from Table 3 follow equally from this table. The range of M being from 10^{-10} to 10^{-22} , the conditions which ensure the

formation of large drops are such as might be expected to arise at times, but not as a matter of course.

It is evident that, as stated in the Introduction, significant growth beyond $r = 100\mu$ requires most unusual circumstances when there are two categories only of nuclei and the motion is sinusoidal. This must also be the case in general, for, apart from the case where the total number of nuclei is small, that with two categories is the most favourable which can be imagined. Also, sinusoidal motion is of the right general character to take advantage of the possibilities for large drop production which are inherent in such a nucleus spectrum.

Two cases can now be distinguished. When $\frac{1}{2} \leq S_1 \leq S_2$ and Table 4 applies, the sizes of the M_1 nuclei are of no moment. Growth occurs preferentially on the M_1 nuclei because they alone are activated; large drops can be formed with comparatively small values of P , provided M_2 is small enough (S_2 large). The growth of the M_1 nuclei depends on a high ambient supersaturation and not on the nucleus they contain. It will be seen from Table 4 that the condition (2.3) is fulfilled over the greater part of the ranges of P and r shown. Where it is not, it is necessary to postulate (2.2).

On the other hand, if M_1 is very large and r_0 not too great, $\frac{1}{2}$ can assume values comparable with some of those in Table 3. Then, even with $S_2 \approx S_1$ quite small, fairly large drops can form with moderate values of P , provided $\frac{1}{2}$ (and hence M_1) is large enough. In this case the growth occurs because of the strong influence of the M_1 nucleus and independent of any ambient supersaturation. The M_1 drops never even reach their critical size. For with $M_1 = 10^{-10}$, $S_1 \approx 10^{-5}$ C. and S_2 need only be larger in the ratio $1.6n_2/n_1^{1/3} : 1$, as required by (2.2). Of course, if M_2 is small (and S_2 large), the growth can be all the more rapid.

These results may be compared with Figure 5 of Part I, where the same phenomena appear. Of course where M_1 is not very large ($\frac{1}{2} = 0$), in sinusoidal motion, S is by no means always equal to S_2 , so that the growths found from this figure for $S = S_2$ over half a period naturally exceed those found here. Also with very large nuclei the influence of the nucleus during growth through the smaller values of r is greater than is allowed for by the term $\frac{1}{2}$, which represents only the minimum effect of the nucleus occurring when $r = r_m$.

IV. CONTINUOUS SPECTRUM OF NUCLEI

(a) Any Motion

Tables 3 and 4 give not only a quantitative answer to the idealized problem with only two categories of nuclei, but also a fairly clear indication of what is possible in general.

In order to show that the result found with only two categories of nuclei does not depend essentially on this assumption regarding the nuclei, the treatment can be extended to take account of continuous spectra. The final result is a method of finding a lower bound for the radius of the largest drops formed in a sinusoidal motion, the nucleus spectrum being arbitrary.

As mentioned in Section II (*b*) above, in a motion obeying the condition (2.7) the n_2 nuclei, of size M_2 , may be replaced by a continuous group of nuclei the largest of which is of size M_2 , without changing anything. The n_1 nuclei, M_1 , may also be replaced by a continuous group containing the same number of nuclei, the smallest of which is of size M_1 . A nucleus larger than M_1 will have a larger initial radius than an M_1 nucleus, whatever the previous history of the air parcel. Hence at any stage during the uplift it will have lower equilibrium vapour pressures and will grow (in mass) more rapidly than an M_1 droplet. It will therefore be more effective than the corresponding M_1 droplet in keeping S less than S_2 , and *some* of them will produce drops of radius exceeding r_f at any level.

The interval between M_1 and M_2 cannot be dealt with so easily. Obviously nuclei of intermediate size cannot be introduced without destroying the argument. If the interval is made small, $(S_2 - S_1)$ becomes small, and (2.7) becomes increasingly stringent, requiring smaller and smaller velocities. A sufficient condition may still be found, however, for the production of drops of a known size. Let $N_1 = N(S_1)$ be the number of nuclei in a continuous spectrum whose critical supersaturations are less than S_1 , that is, whose M values exceed M_1 ; similarly let $N_2 = N(S_2)$. There will be $N_2 - N_1$ nuclei of sizes intermediate between M_1 and M_2 . Suppose now that these intermediate nuclei are removed and in their place an equal number of M_2 nuclei are substituted. Suppose also that the N_1 nuclei larger than M_1 are replaced by an equal number of size M_1 . We have then essentially a two-category spectrum. Provided initially

$$q_1 z_0 > S_2 > S, \dots\dots\dots (4.1)$$

and one of the conditions (2.2) or (2.3) is met, if, in the subsequent motion,

$$v < 4\pi E q_2 N_1 (S_2 - S_1 + \psi) r_f / q_1, \dots\dots\dots (4.2)$$

where

$$r_f^3 = 3(q_1 z - S_2) / 4\pi \rho' q_2 N_2$$

at each level, only the M_1 nuclei will be activated. Further, they will at any level form drops of radius at least r_f . Suppose that such a motion occurs. Let the intermediate nuclei be restored at the commencement of the motion. Some of these may grow beyond the critical stage, some may not. But since they are larger than the M_2 nuclei which they replaced, they will absorb more liquid water, and hence S will remain less than S_2 , *a fortiori*. On the other hand, since they are smaller than M_1 , they will share the liquid water less than proportionally with the M_1 drops. These latter will therefore have radii exceeding $(N_1/N_2)^{1/3} r_f$ at any level. Finally, let the continuous group of N_1 nuclei larger than M_1 replace the N_1 nuclei of size M_1 . As before, S will remain less than S_2 . Further, *some* nuclei of the group will form drops of radius exceeding $(N_1/N_2)^{1/3} r_f$.

Suppose that a continuous spectrum of nuclei is specified, and any motion is defined starting from an initial level z_0 . If it is assumed that the initial condition (4.1) is fulfilled, M_1 and M_2 may be chosen in such a way that (4.2) also holds. In general, this choice may be made in an infinite number of ways,

each of which will yield a rough lower bound for the radius of the largest drops formed at any level. We are at liberty to make the choice which gives the largest value for this bound. It is of course necessary to ensure that (2.2) or (2.3) is fulfilled. Table 4 suggests that, when $\psi = 0$, (2.3) will be found to hold under most conditions. As before, a quantitative result can be obtained only after the form of the motion is specified.

(b) *Sinusoidal Motion with a Continuous Spectrum of Nuclei*

The ideas of the preceding section may now be applied to sinusoidal motion. Suppose that such a motion is given, specified by A and P . By (3.1), the condition (4.2) will be satisfied provided the choice of M_1 and M_2 is made so that

$$\frac{97 \cdot 54}{N_1^{2/3}(S_2 - S_1 + \psi)} \leq \frac{P}{A^{2/3}}, \quad \dots \dots \dots (4.3)$$

If this is done, it follows from the last section that S will never exceed S_2 , and that *some* drops will be formed whose radii exceed $(N_1 N_2)^{1/3} r_c$ at any level; in particular at the apex, some will have radii exceeding r_a where

$$r_a^3 = 3\{q_1(2A + z_0) + q_2(W_0 - w_2) - S_a\}/4\pi\phi'q_2N_2,$$

where S_a is the supersaturation which would exist at the apex if the $(N_2 - N_1)$ nuclei intermediate between M_1 and M_2 were replaced by nuclei of size M_2 and if the N_1 nuclei larger than M_1 had been replaced by nuclei of size M_1 . As is shown in Appendix I, S_a is negligible compared with $2Aq_1$, provided $A > 2 \times 10^4$ cm., and if this is so we may write

$$r_b^3 = 3(2A + z_0)q_1/4\pi\phi'q_2N_2, \\ r_a = r_b, \quad \dots \dots \dots (4.4)$$

whatever the initial radius of these drops.

The choice of S_1 and S_2 is still to some extent arbitrary, and we are at liberty to choose those values which while satisfying (4.3) yield a maximum value for r_b . From (4.4) this will occur with a minimum N_2 . Since N is a monotone increasing function of S (the critical value of S corresponding to each value of M) this will correspond to a minimum of S_2 . Now, with any choice of S_1 , the smallest permissible value of S_2 is that which gives

$$\frac{97 \cdot 54}{N_1^{2/3}(S_2 - S_1 + \psi)} = \frac{P}{A^{2/3}}, \quad \dots \dots \dots (4.5)$$

This may be regarded as defining S_2 in terms of S_1 , so that there is still one degree of freedom left in the choice of M_1 and M_2 . Differentiating (4.5) to find S_1 for a stationary S_2

$$\left[N^{5/3} \frac{d(\psi - S_c)}{dN} \right]_{S_c=S_1} = -\frac{2}{3}C, \quad \dots \dots \dots (4.6)$$

where $C = 97 \cdot 54 A^{2/3}/P$.

This relationship will in general be satisfied at only a finite number of points on the cumulative distribution curve of the given nucleus spectrum.

Equation (4.6) yields in principle the solution of the problem, giving the values of S_1 and N_1 which on substitution in (4.5) give the smallest value for S_2 and hence for N_2 also. But a simple general treatment is possible only for the case $\psi=0$ corresponding to Table 4 above.

For the case where $\psi \neq 0$, the best physical insight is given by Table 3 above. Only the case $\psi=0$ will be considered. The lower bound, r_b , found for r_a is of course valid even when $\psi \neq 0$, but it may be substantially too small.

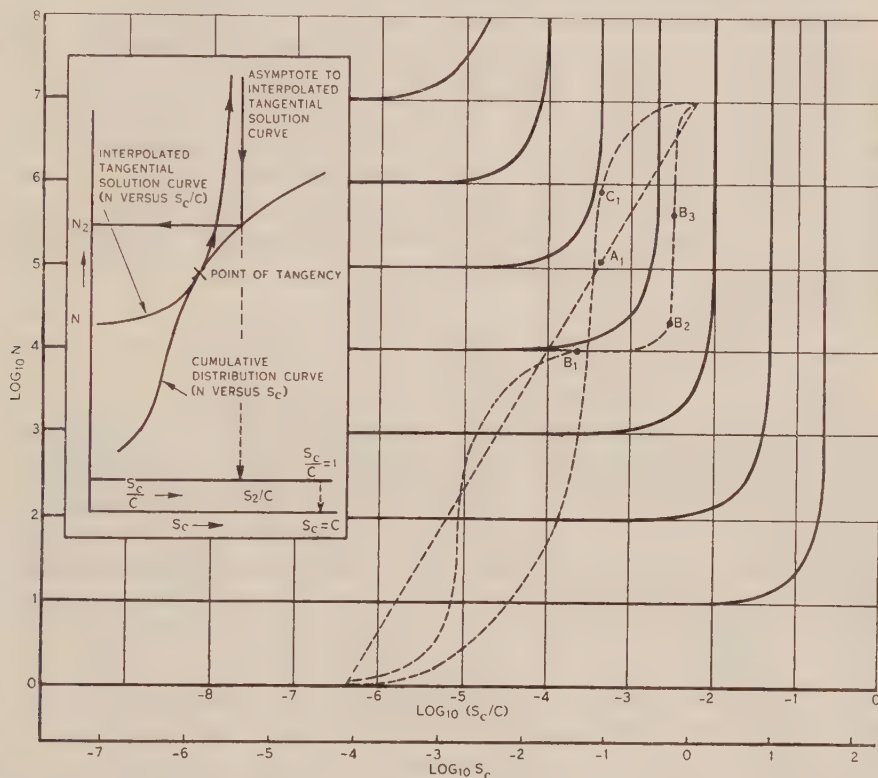


Fig. 1.—Graphical method for finding a lower bound for the radius of the largest drops formed in a sinusoidal vertical motion (continuous spectrum of nuclei).

- Solution curves of (4.6) for evaluating N_2 graphically when $\psi=0$.
- - - - Three cumulative distribution curves drawn arbitrarily for a sinusoidal motion of amplitude 1 km., period 1250 seconds with 10^7 nuclei per gram, their critical supersaturations ranging from 1 to 10^{-4} °C.

The solution curves of (4.6) in the case $\psi=0$ are

$$S_c = Cc_1 - CN^{-2/3}, \quad \dots\dots\dots (4.7)$$

where c_1 is the parameter of the family.

The roots of (4.6) for a particular distribution curve will be given by the points of tangency with this family of solution curves; they give the values of S_1 at which S_2 is stationary.

Since for $\psi=0$, $N_1^{2/3}(S_2-S_1)$ is constant by (4.5), S_2 must have positive infinities at both $S_1=0$, $S_1=\infty$; for other values of S_1 , S_2 is finite. Hence at least one of the roots of (4.6) must correspond to a minimum of S_2 ; if there are several roots, that one will be chosen which gives the smallest value for S_2 . From (4.5) and (4.7), S_2 is given by

$$S_2=S_1+CN_1^{-2/3}=Cc_1, \quad \dots\dots\dots (4.8)$$

where c_1 is the parameter of the tangential solution curve (4.7).

If therefore any distribution of nuclei and any sinusoidal motion are given, the most favourable choice of S_1 and S_2 can be found and from this the best value of r_b . Any motion in which the upward velocity is less, level for level, will produce drops at least as big. From the way in which the rough lower bound is derived, it is obvious that the bound will be fairly exact if the nucleus spectrum consists of two peaks; with the other spectrum forms it will still be valid, but may be only a very rough bound. The method described could be illustrated by assuming some spectrum, defined analytically. A better insight is gained, however, by means of a graphical method of finding the roots of (4.6) for any particular spectrum.

(c) *A Graphical Method for Finding r_b ($\psi=0$)*

When $\psi=0$, the minimum value of N_2 corresponding to one of the roots of (4.6) may be found graphically by means of Figure 1. The dashed curves are three cumulative distribution curves, arbitrarily drawn, relating N to the critical supersaturation S_c , on the bottom scale of the figure. The heavy curves are the family (4.7) relating N to $S_c C$. The two scales of $\log_{10} S_c C$ and $\log_{10} S_c$, and with them the two sets of curves, are brought into correct relationship by adjusting them so that the zero graduation on the $\log_{10} S_c C$ scale corresponds to the C graduation ($C=97.54A^{2.3}P$) on the S_c scale. The N scale is common.

The use of these curves is most easily seen from the key diagram. A solution curve is interpolated which is tangential to the cumulative distribution curve. The point of tangency is a solution of (4.6), giving N_1 and S_1 . From (4.7) it is easily seen that the vertical asymptote to the solution curve has an abscissa equal to Cc_1 , that is to S_2 by (4.8). Its intersection with the distribution curve gives N_2 and hence, r_b . The process of finding N_2 is indicated on the key diagram by arrows.

The points of tangency of the three distribution curves in the main figure are marked A_1 ; B_1 , B_2 , and B_3 ; C_1 . The three distribution curves on Figure 1 all represent spectra in which there are 10^7 nuclei per gram of air, the range of S_c being 1.0 to 10^{-4} °C. The value of C was taken as 168, which corresponds to $A=10^5$ cm., $P=1250$ seconds. The "A" curve represents a power law, the "B" curve a very favourable case where the nuclei occur in two main groups, and the "C" curve a spectrum with a single maximum (on the log-log scale used here).

The approximate values of N_2 corresponding to the points A_1 , B_1 , and C_1 are given in Table 5. The points B_2 and B_3 yield larger values of N_2 than B_1 and therefore are not used. The ratio N_∞/N_2 , where N_∞ is the total number of

nuclei, is the ratio of the mass of a drop of radius r_b to the mass of those which would result if all the nuclei formed equal drops. It may therefore be taken as a measure of the degree to which the heterogeneity of the nucleus spectrum is reflected in the droplet spectrum.

TABLE 5
VALUES OF N_∞/N_2 AND r_b FOR THE THREE DISTRIBUTION CURVES IN FIGURE 1
 $A=10^5$ cm., $P=1250$ sec.

Curve	N_2	N_∞/N_2	r_b (μ)
"A"	5.0×10^5	20	12
"B"	1.6×10^4	625	39
"C"	2.0×10^6	5	8
All nuclei same size	10^7	1	6

It is easily verified that the solution for the "B" curve satisfies the condition (2.3).

These numerical results are purely illustrative and cannot be taken as giving anything more than a very rough idea of the drop sizes that may result from condensation. They are merely intended to show how critically the form of the nucleus spectrum (as distinct from the total number of nuclei) might influence the outcome of one particular motion. Examples can easily be constructed to illustrate the point for other sinusoidal motions; it is merely necessary to transpose the distribution curves horizontally to take account of the new value of C . As C decreases, these curves and the $\log_{10} S_c$ scale move to the right, and it is evident that this results in smaller values of N_2 . It follows that for a given distribution and a fixed value of A , r_b increases with P . However, the relation of r_b to A for fixed P is not so simple. As A increases, so does C , and hence N_2 , so that although more liquid water is released it may have to be shared by more drops.

A few trials on Figure 1 soon make it obvious that large values of r_b (that is, small values of N_2) can result only if C is not too large. Also, for the particular value of C used here, the nucleus spectrum must be such that only a few of the nuclei have S_c values smaller than about 0.1°C ., that is M values greater than about 10^{-18} . This is in accordance with the general impression given by Table 4.

(d) The Total Number of Nuclei

Our knowledge of the concentration of nuclei in various air masses is based almost wholly on observations with the Aitken counter, in which the air is expanded more or less adiabatically by 20 per cent. in volume. Air in the central part of the chamber will undergo a practically adiabatic change, giving a cooling of about 20°C ., and a supersaturation, S , of about 16°C . Since successive expansions are made until no more nuclei fall out, this instrument

must count all nuclei down to $S_c = 16^\circ\text{C}$., $M \approx 5 \cdot 10^{-23}$. It is well known that the concentration of drops in cloud is much smaller than that commonly found for nuclei. Only the larger of the Aitken nuclei can normally become operative in cloud.

Landsberg (1938) gives the average concentration of nuclei over the ocean (600 observations at 21 places) as 940 per cc.: inland, in the country, as 9500 (25 places, 3500 observations). In towns and cities the average counts are of course much bigger. If we assume for the moment that the shape of the cumulative distribution curve is the same over land and sea it is easily seen from (4.6) that a tenfold reduction in N for each value of S_c is equivalent to decreasing C by a factor of $10^{2/3} \approx 4.6$. This implies that over the sea a period of $P = 1000$ seconds would (on the average) produce the same value of N_2 and hence of r_b as a period of 4600 seconds over the land: alternatively, with a given value of P , A could be increased 10 times over the sea, yielding the same value of N_2 , with r_b increased by a factor of $10^{1/3} \approx 2.2$.

The assumption that the cumulative distribution curves are the same shape over land and sea is of course untrue. On the average, nuclei would be larger over the sea. At least in the case where $\psi = 0$, this operates to reduce the advantage of maritime air masses. On the other hand, maritime air must often contain less than 940 nuclei per cc.: clouds in such an air mass could be very apt to produce large drops by condensation alone.

(e) *The Effect of Temperature and of Sedimentation*

In Section IV (f) of Part I it was shown that ξ_c changes only slightly with temperature. Hence for given values of M_1 and M_2 , S_1 and S_2 change only slightly. On the other hand, E increases almost by a factor of two between 0 and 30°C .

In the case of two categories of nuclei, equation (3.5) shows that the factor E takes account of the influence of temperature on the process we are considering. With all other variables constant, the lower bound on P is inversely proportional to E , and so has to be nearly twice as great at 0°C . as at 30°C .

With a continuous spectrum the relationship is not quite so clear cut. The numerical constant occurring in equation (4.5) is proportional to $1/E(q_1 q_2)^{2/3}$, as is seen from (3.1). Hence if P is decreased in proportion to $E(q_2 q_1)^{2/3}$, (about 25 per cent. between 0 and 30°C .) C remains unchanged, and so must the resulting values of S_1 and N_2 . On the other hand, r_b is proportional to $(q_1 q_2)^{1/3}$ for fixed A , z_0 , and N_2 ; that is, it is about 22 per cent. larger at 30°C . than at 0°C .

Hence in all cases increasing temperature favours the formation of large drops. Kraus and Smith (1949) found that larger drops resulted at higher temperatures. But Tables 3 and 4, or an inspection of Figure 1, show that, in nature, the variation of P and of the nucleus spectrum is far more significant, unless indeed the ranges of such parameters as P and M_2 shown in Table 4, for example, are much wider than they should be.

With the long period motions envisaged here, the vertical velocity will often be small, and it is evident that sedimentation might become important.

It is not proposed to go into this question, but it can be seen that the first order effect would be as follows: a drop which lags behind the air motion near the apex would experience a higher supersaturation than one which did not, for, towards the apex, S is becoming smaller with decreasing v ; thus, when the air reached the apex, the drop would be some way below it and it would be larger than if sedimentation had not occurred.

The neglect of sedimentation in the calculations made does not seem likely therefore to invalidate them, for we have computed, in Tables 3 and 4, *sufficient* conditions for the formation of drops of a certain size, and in Figure 1, a *lower* bound for the drops produced in a given sinusoidal motion.

V. COLLOIDAL INSTABILITY IN NON-FREEZING CLOUD

(a) *The Significance of the Condensation Phase*

The view of rain from non-freezing cloud suggested by the considerations of the Introduction may be expressed as follows. The observed stability of most water clouds in temperate latitudes suggests that there is a gap between the sizes normally reached as a result of condensation and that size at which coalescence begins to be effective. It seems possible that normal cloud drops of order $r=10\mu$ do not grow significantly by coalescence in the times commonly available either because such small drops come into contact too rarely, or because contact does not result often enough in coalescence.

If then a water cloud is to form rain, this gap between the condensation and the coalescence phases must be bridged. This can happen either because larger drops are formed by condensation or because coalescence begins to operate at smaller sizes. It is difficult to see what influence could cause the efficiency of the coalescence process to vary from one cloud to another, bearing in mind that in respect of depth, temperature, and motion, unstable clouds do not seem very different from many stable ones. We are left, therefore, with the alternative that certain circumstances arise at times, particularly in tropical maritime air, which favour the formation of larger drops than usual at least in some parts of the cloud during the condensation phase. If this theory is sound, these circumstances cannot be normal, at any rate in temperate latitudes.

(b) *The Formation of Large Drops by Condensation*

The analysis given above indicates that very large cloud drops can be formed in a parcel of cloud air moving upwards in a sinusoidal motion with a period of the order of one thousand to several thousand seconds if the nucleus spectrum is one of two types:

- (i) The spectrum includes a few "giant" nuclei ($M=10^{-10}$ to 10^{-12}), the total number of such nuclei not exceeding some tens per cc. In this case, large drops can form on the "giant" nuclei without any appreciable supersaturation. The remaining nuclei must be appreciably smaller, as required by equation (2.2).

- (ii) Only a few nuclei (some tens per cc.) are larger than a certain limit which varies as P^2 and is of the order of $M=10^{-20}$ to 10^{-16} for $P=500$ – $10,000$ seconds. In this case the supersaturation can be appreciable (of order 10^{-2} to 1 °C., ≈ 0.06 to 6 per cent. R.H.) without activating the remaining smaller nuclei, and the few larger ones can grow accordingly, even though their size does not greatly exceed the limit mentioned, and the nucleus itself contributes little to their speed of growth.

(c) *The Motion of the Air Parcel*

As pointed out previously, the motion need not be accurately sinusoidal. Any motion where an air parcel starts from a lower level somewhere in the cloud, moves upward and finally comes to rest (or nearly so) at some upper level may be regarded as a sinusoidal motion, specified by an amplitude and a period. Since the motion considered is special to the parcel and does not involve the cloud as a whole, it seems likely *a priori* that in most clouds where the vertical velocities are not too great some parcels will at times move in the way required, that is with effective values of P of the order of some thousands of seconds. This cannot, of course, be established by observation.

However, Workman and Reynolds (1949) found that cumulus tops rise and fall with a period of about 17 minutes in the case of clouds which returned a radar echo and 13 minutes in those which did not. Smith (1951, flight on March 24, 1950) describes how several humps at the top of a raining water cloud rose and fell with a period of the order of 10 to 15 minutes. These observations may perhaps be taken as indicating the periods of vertical oscillations which commonly occur in cumuliform clouds. Longer period oscillations, even if they involve portions of the cloud large enough to affect its visible outline, would naturally be harder to observe and would be masked more easily by other developments.

The average vertical velocity in the postulated sinusoidal motion gives the best check available at present on the required values of the period P . This will be $2A/P$; for $A=2$ km., $P=10^4$ sec., this is 0.4 m. per sec., so that even the longest periods shown in Tables 3 and 4 are not unreasonable. The conclusion that low average vertical velocities favour colloidal instability is consistent with most of the observations on turbulence in raining water clouds (Kotsch 1947; Smith 1951).

(d) *The Spectrum of the Condensation Nuclei*

Unlike the motion, which is special to the parcel considered, the nucleus spectrum is presumably characteristic of the air mass in which the cloud forms. It is therefore, in principle, easier to observe. However, at present, little is known of nucleus spectra beyond the range of M , from about 10^{-10} to 10^{-22} (S_c about 10^{-5} to 10^{-8} °C.). The only measurements of the spectrum seem to be those of Woodcock and Gifford (1949) at the larger end, and those of Junge (1935) at the smaller. The latter were made in surface air at Frankfurt, and are therefore not very relevant. The requirements on the nucleus spectrum for

the formation of large drops are most clearly shown by Tables 3 and 4. In view of the range of S_c and M , these would appear to be possible, but not too easy, to fulfil; for all but a few (some tens per cc.) of the nuclei must be smaller than M_2 if (3.4) is to be satisfied by reasonable values of $(2A+z_0)$ for the large values of r_b quoted. This gives some confirmation to the view of rain from non-freezing cloud stated in Section V (a) above. So also does the discussion of Section IV (d), which affords some explanation of the fact that such rain has mostly been observed over or near the sea.

The view that rain from non-freezing cloud is due to the presence of very large sea-salt nuclei receives support from the observations of Woodcock and Gifford (1949) and Woodcock (1950). Nuclei in the "giant" category (10^{-10} to 10^{-12}) were found up to cloud levels. These measurements were taken over the sea in the Bermuda area in fine weather with a surface relative humidity of 42 per cent. and two-tenths cumulus, and off Miami with surface relative humidities ranging from 68 to 82 per cent. Now the falling velocity of nuclei of a given size depends on the relative humidity. Their chance of reaching cloud levels depends also on turbulent diffusivity which itself depends on the stability of the air, wind strength, and the nature of the surface. With so many variables entering, it is by no means certain whether these observations can be taken as representing the nucleus spectrum of air reaching the condensation level in rainy situations. Again, the effect of rain itself must be considered. Landsberg (1938) cites several sets of observations showing that the total number of nuclei decreases during rain. Obviously the larger nuclei would tend to be captured more. Consider nuclei $M=10^{-12}$, 10^{-10} in air where $S=-1^\circ\text{C}$. (≈ 94 per cent. R.H.). They have equilibrium radii of 4μ and 18μ respectively. Their growth is mainly controlled by the fifth-power law of equation (4.8) of Part I, up to fairly near their equilibrium size. It is easily seen from this that they approach their equilibrium sizes fairly nearly in a few tens of seconds even starting from zero size. They may therefore be assumed to be in equilibrium. According to a table given by Langmuir (1948) a 1 mm. diameter raindrop will capture about 50 and 90 per cent. respectively of the droplets of the above sizes that lie in its path. A few millimetres of rain composed of such drops would evidently reduce drastically the numbers of such nuclei. On the other hand, such a raindrop would not capture any droplets at all smaller than about $r=1\mu$; if $M<2.6\times 10^{-16}$, even at saturation the nucleus will not form a droplet larger than this. These results are only slightly different for raindrops of other sizes. Hence on the basis of Langmuir's table, and assuming that contact implies coalescence, at least in a significant proportion of cases, rain provides a very effective means of removing these "giant" nuclei larger than $M=10^{-12}$, while those smaller than $M=10^{-16}$ are unaffected by it, at any rate until they have entered the cloud and grown to sizes beyond $r=1\mu$ in supersaturated air.

In cases where rain is widespread and long continued, the necessary supply of water vapour must come in with air which enters the cloud from below. The air will therefore be rainwashed, and probably devoid of "giant" nuclei by the

time it reaches the cloud base. In such cases, therefore, the nucleus spectrum of the condensing air could not conform to type (i). On the contrary, the removal of the larger nuclei would tend to produce one of type (ii). A situation of this kind would contain a regenerative element, rain itself producing conditions favouring the formation of still more rain.

This effect might be expected to enter, even when raindrops originate as ice crystals. As mentioned above, the evidence of chlorine content measurements on cloud and rain (Wright 1940) indicates that they grow largely by coalescence, and it seems likely, according to the table of collection efficiencies given by Langmuir (1948), that such growth will be enhanced if the cloud contains some larger cloud drops than are usual.

Such a mechanism might go some way towards explaining the impression formed by some meteorologists that rainy or dry periods tend to persist and are broken only by a comparatively large change in the situation.

VI. CONCLUSION

The hypothesis advanced, that colloidal stability and instability in water clouds is largely determined by the size of the largest drops resulting from condensation, has proved satisfactory in as much as the conditions for the formation of large drops in some air parcels appear to be possible but probably not frequent. Further, it is consistent with the predominantly maritime occurrence of rain from non-freezing clouds.

The vertical motion must be relatively slow, particularly in the initial stages. This condition seems likely to be met by some air parcels in many clouds. The nucleus spectrum must be rather unusual, containing either a few "giant" nuclei (tens per cc.) or else a similar number of large or moderately large nuclei, with all the rest smaller than a certain limit that depends on the motion. This is probably the true limiting factor; it seems likely to be met at times but not in general. Moreover, as it is characteristic of the air mass, it is more easily observed than the motion of an individual air parcel.

The chief result of this investigation is to point to the need for observations of the spectrum of condensation nuclei, about which practically nothing is known. Such observations are needed both in air masses where water clouds of adequate depth remain stable and in those where they yield rain.

VII. LIST OF SYMBOLS

- A , amplitude of a sinusoidal motion,
- C , see equation (4.6),
- M_1, M_2 , the sizes of two groups of nuclei (\approx molar mass),
- N_1, N_2 , number of nuclei with critical supersaturations less than S_1, S_2 ,
- N_∞ , total number of nuclei,
- P , period of a sinusoidal vertical motion,
- S_0 , the value of S at $t=0, z=z_0+A$ (Appendix I),
- S_1, S_2 , critical supersaturation of a nucleus of size M_1, M_2 ,

- S_a , the value of the supersaturation, S , at the apex of a sinusoidal motion,
 S_c , the critical supersaturation of a nucleus ($=\xi_c M^{-1}$),
 W_0 , the total liquid water associated with the M_1 and M_2 nuclei at $S=0$,
 Y , a variable which dominates y (Appendix I),
 Y_0, Y_a , the values of Y at $z=z_0+A, z_0+2A$ (Appendix I),
 c_1, c_1' , values of constant of integration in equation (4.7),
 $f_1 = (q_1 z_0 - S_2)/q_1 A$ (Section III (a)),
 $f_2 = 0.916 (S_2 - S_1 + \psi)/\{1 - 0.0068 (A q_1)^{-4/3}\}$ (Appendix I),
 l , the solution of

$$\exp\left(\frac{\pi A q_1 (1 - 0.0068 (A q_1)^{4/3})}{2 \times 0.916 \times 1.8}\right) = 2 + \frac{1}{A q_1} \quad (\text{Appendix I}),$$
 n_1, n_2 , the number of nuclei of sizes M_1 and M_2 ,
 r , radius of the M_1 droplets,
 r_a , radius of the M_1 droplets at the apex of a sinusoidal motion,
 r_b , a lower bound for r_a ,
 r_c , the critical radius of a drop ($=\eta_c M^{1/2}$),
 r_f , see equation (2.5),
 r_0 , the radius of a drop at $t=0, z=z_0+A$ (Appendix I),
 r_m , the maximum drop radius considered, see equation (2.4),
 v_f , see equation (2.6),
 w_1, w_2 , the mass of liquid water associated with the M_1 and M_2 nuclei,
 $x = A q_1 / f_2$ (Appendix I),
 $y = S - \beta/r + \gamma M_1 / r^3$ (Appendix I),
 y_a , the value of y at the apex of a sinusoidal motion (Appendix I),
 y_0 , the value of y at $t=0, z=z_0+A$ (Appendix I),
 z_0 , the height of the initial level above the condensation level,
 $\zeta = (z - z_0)/A$ (Section III (a)),
 ψ , the lower limit of $S_1 - \beta/r + \gamma M_1 / r^3$, see equation (2.4), for $r < r_m$.

VIII. REFERENCES

- BYERS, H. R. (1949).—*J. Met.* **6**: 363.
 CRADDOCK, J. M. (1949).—*Quart. J. R. Met. Soc.* **75**: 147-53.
 DAVIES, D. A. (1950).—*Met. Mag.* **79**: 354.
 FOSTER, H. (1950).—*Bull. Amer. Met. Soc.* **31**: 140-1.
 HEYWOOD, G. S. P. (1940).—*Quart. J. R. Met. Soc.* **66**: 46.
 HOWELL, W. E. (1949).—*J. Met.* **6**: 134-49.
 HUNT, T. C. (1949).—*Met. Mag.* **78**: 26.
 JUNGE, C. (1935).—*Beitr. Geophys.* **46**: 108-29.
 KINSHOTT, E. J. (1950).—*Met. Mag.* **79**: 265-6.
 KOTSCH, W. J. (1947).—*Bull. Amer. Met. Soc.* **28**: 87-9.
 KRAUS, E. B. (1949).—*Met. Mag.* **78**: 232.
 KRAUS, E. B., and SMITH, BETTY (1949).—*Aust. J. Sci. Res. A* **2**: 376-88.
 LANDSBERG, H. (1938).—*Ergebn. Kosm. Phys.* **3**: 155-252 (p. 200).
 LANGMUIR, I. (1948).—*J. Met.* **5**: 175-92.

- LEOPOLD, L. B., and HALSTEAD, M. H. (1948).—*Bull. Amer. Met. Soc.* **29**: 525-34.
 SMITH, E. J. (1951).—*Quart. J. R. Met. Soc.* **77**: 33-43.
 SQUIRES, P. (1952).—*Aust. J. Sci. Res. A* **5**: 59.
 SWINBANK, W. C. (1947).—*Nature* **159**: 849-50.
 VIRGO, S. E. (1950).—*Met. Mag.* **79**: 237-8.
 WEXLER, H. (1945).—*Bull. Amer. Met. Soc.* **26**: 156-9.
 WOODCOCK, A. H. (1950).—*J. Met.* **7**: 161-2.
 WOODCOCK, A. H., and GIFFORD, MARY M. (1949).—*J. Mar. Res.* **8**: 177-97.
 WOOD JONES, F. (1934).—"Unscientific Excursions." p. 124. (Edward Arnold and Co.: London.)
 WORKMAN, E. J., and REYNOLDS, S. E. (1949).—*Bull. Amer. Met. Soc.* **30**: 359-61.
 WRIGHT, H. L. (1940).—*Quart. J. R. Met. Soc.* **66**: 3-12.

APPENDIX I

The Behaviour of S near the Apex of a Sinusoidal Motion

In order to find an upper limit for S_a which we neglected in equation (3.4), assume for the moment that the M_2 drops may be disregarded. It is convenient to introduce a new variable

$$y = S - \frac{\beta}{r} + \frac{\gamma M_1}{r^3}.$$

Then

$$\begin{aligned} \frac{dy}{dt} &= \frac{dS}{dt} + \frac{\beta}{r^2} \left(1 - \frac{r_c^2}{r^2} \right) \frac{dr}{dt} \\ &= q_1 v - 4\pi E F(R_c, \sigma') r y \left\{ n_1 q_2 - \frac{\beta}{4\pi \rho' r^4} \left(1 - \frac{r_c^2}{r^2} \right) \right\}, \quad \dots \quad (\text{A1}) \end{aligned}$$

by (4.2) and (2.6), Part I, ($r_c = \sqrt{3\gamma M_1 \beta}$). This equation will be applied over the interval $t=0$ to $t=P/4$ ($z=z_0+1$ to $z_0+2.1$) to find an upper bound for y_a at $t=P/4$ in terms of y_0 (at $t=0$). In this interval

$$r^3 > 3q_1 A / 4\pi \rho' q_2 n_1, \quad \dots \quad (\text{A2})$$

by (2.8).

The second term in the bracket may be dealt with as follows:

$$\frac{\beta}{4\pi \rho' r^4} \left(1 - \frac{r_c^2}{r^2} \right) = \frac{\beta}{4\pi \rho' r^4} - \frac{\beta (4\pi \rho' n_1 q_2)^{1/3}}{(3q_1 A)^{4/3}} n_1 q_2.$$

Hence, since n_1 is certainly less than 10^8 ,

$$n_1 q_2 - \frac{\beta}{4\pi \rho' r^4} \left(1 - \frac{r_c^2}{r^2} \right) > n_1 q_2 [1 - 0.0068 (A q_1)^{-4/3}].$$

This will be supposed positive; as it will be, provided $A q_1 > 0.024$ ($A > 3$ m.). Further, $F(R_c, \sigma') > 1$. Suppose first that the value of y at $t=0$, y_0 , is positive. Then since r is positive, it follows from (A1) that $y > 0$ from $t=0$ to $P/4$. Hence

$$\frac{dy}{dt} < q_1 v - 4\pi E q_2 n_1 r y [1 - 0.0068 (A q_1)^{-4/3}].$$

By (A2) above and (3.1)

$$4\pi E q_2 n_1 r > 2\pi A q_1 / 0.916 P (S_2 - S_1 + \psi).$$

Hence

$$\frac{dy}{dt} < \frac{2\pi A q_1}{P} \left(\cos \frac{2\pi t}{P} - \frac{y}{f_2} \right),$$

where

$$f_2 = 0.916(S_2 - S_1 + \psi) / (1 - 0.0068(Aq_1)^{-4/3}).$$

If $y_0 < 0$ the inequalities used above do not apply. But the solution of (A1) satisfying the condition $y = y_0$ at $t = 0$ is a non-decreasing function of y_0 . For suppose that, for any given v as a function of t , at some value of t two solution curves meet. Then since y has the same value on the two curves, so must r , and hence dy/dt must be the same for them both; that is, they must be tangential, and can never cross.

Hence for all y_0 , the solution of (A1) is dominated by that of

$$\frac{dY}{dt} = \frac{2\pi A q_1}{P} \left(\cos \frac{2\pi t}{P} - \frac{Y}{f_2} \right),$$

satisfying $Y = Y_0$ at $t = 0$, provided Y_0 is positive and greater than y_0 .

Integrating this equation gives the value of Y at the apex, $t = P/4$:

$$\frac{Y_a}{Aq_1} = \left(\frac{Y_0}{Aq_1} - \frac{x}{1+x^2} \right) e^{-\pi x/2} + \frac{1}{1+x^2},$$

where $x = Aq_1/f_2$. Let r_0 be the radius of the M_1 drops at $t = 0$, and let Y_0 be chosen equal to $S_2 - \frac{\beta}{r_0} + \frac{\gamma M_1}{r_0^3}$. Y_0 is positive, for $S_2 > S_1 \geq \frac{\beta}{r} - \frac{\gamma M_1}{r^3}$ for all r . Then, since $S_0 < S_2$, $y_0 < Y_0$. Hence $y_a < Y_a$ at the apex, and

$$\frac{S_a}{Aq_1} \leq \left(\frac{S_2}{Aq_1} - \frac{x}{1+x^2} \right) e^{-\pi x/2} + \frac{1}{1+x^2} + \frac{\beta}{Aq_1} \left(\frac{1}{r_a} - \frac{e^{-\pi x/2}}{r_0} \right) - \frac{\gamma M}{Aq_1} \left(\frac{1}{r_a^3} - \frac{e^{-\pi x/2}}{r_0^3} \right).$$

The positive term $\beta/(Aq_1 r_a)$ is less than 10^{-3} for $r_a > 20\mu$, $Aq_1 > 1$. It will be seen later that this error is negligible, and that for other reasons Aq_1 must be taken greater than unity. Now

$$r_a^3 \leq r_0^3 + \frac{3}{4\pi \rho' q_2 n_1} (S_2 + Aq_1),$$

unless S_a is negative, and

$$r_0^3 \geq 3Aq_1 / 4\pi \rho' q_2 n_1,$$

by (2.8).

Hence

$$r_a^3 / r_0^3 \leq 2 + \frac{S_2}{Aq_1} < 2 + \frac{1}{Aq_1},$$

since S_2 is assumed to be less than unity. Hence the γ term is greater than

$$\frac{\gamma M_1}{Aq_1 r_a^3} \left(1 - \left(2 + \frac{1}{Aq_1} \right) e^{-\pi x/2} \right).$$

Taking the ranges of M_1 and r_m in Table 1 as covering the values of interest, $\psi < 0.80^\circ\text{C}$. Hence the maximum value of $(S_2 - S_1 + \psi)$ is 1.8°C . If this value is substituted in f_2 , this gives a lower bound for x as a function of Aq_1 , and hence also a lower bound for the γ term. Also, f_2 is a decreasing function of Aq_1 , so that x and the bracket

$$\left[1 - \left(2 + \frac{1}{Aq_1}\right)e^{-\pi x^2}\right]$$

are increasing functions of Aq_1 . The γ term will therefore always be positive provided $Aq_1 > l$, where l is the solution of

$$\exp\left\{\frac{\pi Aq_1[1 - 0.0068(Aq_1)^{-4/3}]}{2 \times 0.916 \times 1.8}\right\} = 2 + \frac{1}{Aq_1}.$$

The value of l , found graphically, is 1.116. Since $r_a < r_m$, $\psi < \frac{\gamma M_1}{r_a^3}$. Hence, provided $Aq_1 > l$ ($A > 141$ m.) the γ term is positive and greater than

$$\frac{\psi}{Aq_1} \left[1 - \left(2 + \frac{1}{Aq_1}\right)e^{-\pi x^2/2}\right].$$

Hence

$$\frac{S_a}{Aq_1} < \left(\frac{1}{Aq_1} - \frac{x}{1+x^2}\right)e^{-\pi x^2/2} + \frac{1}{1+x^2} - \frac{\psi}{Aq_1} \left[1 - \left(2 + \frac{1}{Aq_1}\right)e^{-\pi x^2/2}\right].$$

With Aq_1 and ψ constant, this is easily seen to be a decreasing function of x ($x > 0$) by grouping the first two terms in x and differentiating. Hence it is an increasing function of $(S_2 - S_1)$ and it is sufficient to compute it for $S_2 - S_1 = 1$. The resulting values are shown in Table 6.

TABLE 6
UPPER BOUNDS ON S_a/Aq_1

$Aq_1 \backslash \psi$	1.5	2	2.5	3	4	5	7	10	20	50	100
0	0.303	0.178	0.119	0.088	0.050	0.033	0.019	0.009	0.004	0.0005	0.0001
0.4	0.318	0.142	0.072	0.036	0.003	-0.015	-0.025	-0.024	-0.016	-0.008	-0.004
0.8	0.520	0.157	0.060	0.003	-0.047	-0.058	-0.060	-0.054	-0.033	-0.015	-0.008

The greatest values occur with $\psi = 0$ for $Aq_1 \geq 2$; at $Aq_1 = 1.5$, the greatest value occurs at 0.8. Taking the value of ψ which gives the greatest value for the particular value of Aq_1 , the upper bounds as shown in Table 7 result.

The positive β term neglected earlier in the upper bound for S_a/Aq_1 , which was shown to be less than 10^{-3} for $r_a < 20\mu$, $Aq_1 < 1$, is evidently negligible.

It is necessary now to take account of the effect of the M_2 drops. As they are in the stable state, they will tend to make the variation of S more sluggish. An upper limit to the effect they can produce is found by supposing that they

are introduced at the apex and there completely evaporated, the vapour so formed all going to increase the supersaturation. As their radii cannot exceed $\eta_c M_2^{\frac{1}{2}}$, the consequent increase in S could not exceed

$$\frac{4}{3}\pi n_2 \rho' q_2 \eta_c^3 M_2^{3/2}.$$

This is small when M_2 is small; for $M_2 < 10^{-18}$, and $n_2 < 10^8$, it is less than about $2 \cdot 10^{-3}^\circ\text{C}$. This is evidently quite negligible in comparison with the upper

TABLE 7
THE ERRORS CAUSED BY NEGLECTING S_a IN (3.2)

A (m.)	Aq_1	Upper Bounds on :			
		$\frac{S_a}{Aq_1}$	S_a	Proportional Error (%) Due to the Neglect of S_a in :	
				r_a^3	r_a
200	1.5861	0.420	0.666	27.6	8.2
300	2.3791	0.130	0.309	7.0	2.3
400	3.1722	0.076	0.241	4.0	1.3
500	3.9652	0.050	0.198	2.6	0.9
700	5.5514	0.029	0.161	1.5	0.5
1000	7.9304	0.016	0.127	0.8	0.3
2000	15.861	0.005	0.079	0.3	0.1
5000	39.652	0.001	0.040	0.1	0.03

bounds given for S_a in Table 7. On the other hand, if M_2 is large, no demonstration is required to show that S_a is small, for it must be less than S_2 . For $M_2 > 10^{-18}$, $S_a < S_2 < 0.1^\circ\text{C}$. which is smaller than most of the upper bounds given in Table 7.

Thus provided $A > 200$ m., the error in r_a due to neglecting S_a in (3.2) cannot exceed 8 per cent. ; for larger values of A it is much smaller.

TENSAMMETRY : A METHOD OF INVESTIGATING SURFACE PHENOMENA BY A.C. CURRENT MEASUREMENTS

By B. BREYER* and S. HACOBIAN*

[*Manuscript received April 9, 1952*]

Summary

A new type of investigation into surface phenomena, "tensammetry", has been developed by superimposing a small sinusoidal A.C. voltage upon the direct potential applied to a dropping mercury electrode in the presence of surface active substances and measuring the resultant A.C. currents. Wave shaped current-voltage curves are thus obtained which have their origin in adsorption processes occurring at the electrode. These waves are attributed to the movement of surface active molecules in the region of "active space" near the electrode without actual electron transfer across the electrode boundary; that is, the electrode remains polarized with respect to D.C., but is depolarized with respect to A.C. The general properties of tensammetric waves of a number of organic compounds together with their effects on one another and on D.C. and A.C. polarographic waves have been investigated. The theoretical basis and the fundamental equations governing the tensammetric process are derived and discussed.

I. INTRODUCTION

The behaviour of the electrocapillary curve of mercury in the presence of surface active organic compounds was first examined by Gouy (1903, 1906*a*, 1906*b*) who observed that generally the amount of substance adsorbed at the mercury electrode decreased with increasing field strength. Theoretical interpretations of these phenomena were attempted by various workers (e.g. Frumkin 1926; Butler 1929) who derived general equations for the variation of the amount of substance adsorbed with the potential difference at the interface of the electrode. The picture at which these authors arrived was that of the existence of layers of polar molecules (most commonly monolayers) of sparingly soluble organic substances at the electrode boundary (see, for example, Frumkin and Gorodetskaya 1928; Frumkin, Gorodetskaya, and Chugunov 1934; Proskurnin and Frumkin 1935; Gorodetskaya 1940). All the experimental evidence indicates that if the potential of the mercury electrode is made strongly positive or strongly negative, the adsorbed polar molecules at the mercury surface are displaced by the ions of the supporting electrolyte. The displacement frequently occurs over a small range of potential and a sudden change in the charge density of the double layer results. This in turn leads to a large value of the dynamic capacitance ($dq.dE$). Frumkin and Proskurnin (1935) thus observed two maxima with a strong depression in the middle part in the capacity-potential curve of sodium sulphate saturated with octyl alcohol. The extreme anodic and cathodic ends of the curve coincided with those obtained with

* Physico-Chemical Laboratory, Faculty of Agriculture, University of Sydney.

sodium sulphate alone. Grahame (1946) repeated these experiments with similar results. The technique employed consisted in balancing out the double layer capacity with a set of variable capacitors suitably arranged in a Wheatstone network, whilst a small A.C. voltage was superimposed on the applied D.C. potential. The process of balancing was laborious and frequently the peaks of the capacity-voltage curves were not sharply defined.

When electrolytes containing small amounts of surface active substances (e.g. aniline, pyridine, phenol, cresol, benzoic acid, octyl alcohol,* etc.) were subjected to A.C. polarography (Breyer and Gutmann 1946, 1947; Breyer, Gutmann, and Hacopian 1950) current-voltage curves similar to the capacitance-voltage curves of Proskurnin and Frumkin (1935) and of Grahame (1946) were observed by the present authors. But whereas the waves in A.C. polarography are produced as a consequence of depolarization processes occurring at the dropping mercury electrode (in virtue of the applied D.C. potential), the A.C. waves obtained in the presence of surface active substances are produced without corresponding D.C. current flow. In other words: *the electrode is depolarized with respect to A.C. only and remains polarized with respect to D.C.* It is proposed to distinguish this type of electrode process, which is fundamentally different from A.C. polarography, by terming it "tensammetry".

Tensammetric waves are basically an outcome of the same electrode phenomena as those responsible for the capacitance-voltage curves of Frumkin and Grahame. But unlike capacitance-voltage curves, tensammetric current-voltage curves are well defined and the summit potential generally can be read to better than 1 mV. Finally, the time needed to obtain a tensammetric curve is only a fraction of that required for capacitance-voltage measurements.

It is the main object of the present paper to discuss the theoretical basis and to derive the fundamental equations governing the tensammetric process.

II. THE BEHAVIOUR OF DIPOLES AT AN ELECTRODE AND THE GENERAL PROPERTIES OF TENSAMMETRIC WAVES

Consider an orientated dipole in the non-uniform field at the electrode surface (Fig. 1 (a)) and whose charges δ_- and δ_+ are separated by a distance l cm. Let the pole δ_- be placed a cm. from the electrode surface. \vec{E}_1 , \vec{E}_2 , and \vec{E}_3 ($E_1 > E_2 > E_3$) denote the field strengths at the electrode boundary, at a cm. and at $a + l$ cm. from the electrode respectively. Considering the direction normal to the electrode surface only, the resultant force acting on the dipole is represented by

$$(E_2\delta - E_3\delta)\vec{i},$$

where \vec{i} represents a unit vector in the said direction. Since $E_2 \neq E_3$ this force will always be finite and hence the dipole will behave as if it were a charged particle (cf. in this connection, Heyrovsky 1934; Ilkovič 1936; Kolthoff and Lingane 1941). It is proposed to call this type of particle an "active charge". The magnitude and sign of the active charge depends upon the orientation of

* 2-Ethyl hexanol was used throughout.

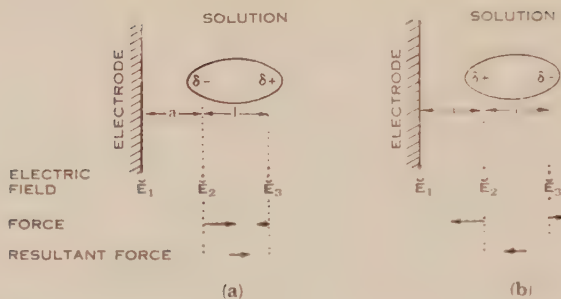


Fig. 1.—The diagrammatic representation of a dipole in a non-uniform electric field close to the electrode-solution interface.

$(E_2 - E_3)\delta$, magnitude of resultant force on dipole; l , length of dipole; a , distance of one end of dipole from the electrode surface; E_1 , field strength at the electrode; E_2 , field strength at distance a from the electrode; E_3 , field strength at distance $a + l$ from the electrode.

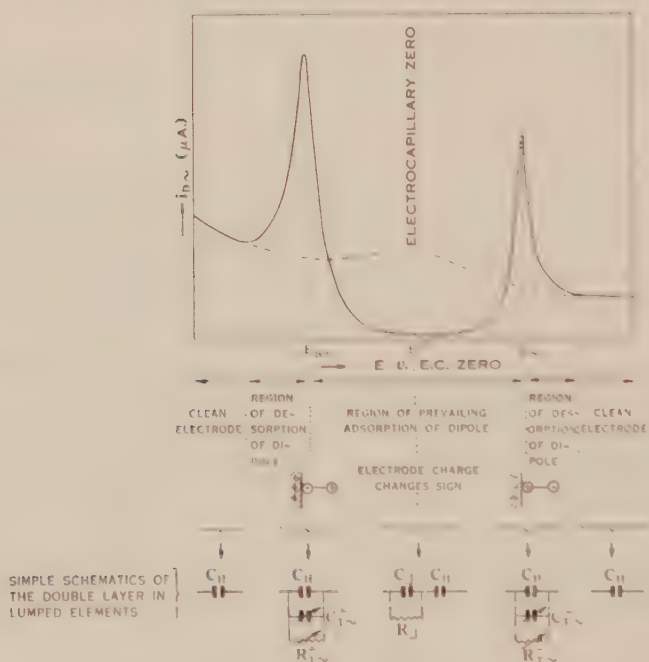


Fig. 2.—Schematic representation of tensammetric waves.

— Tensammetric curve.

- - - Polarogram of supporting electrolyte.

E_{s+} , E_{s-} , summit potentials of positive and negative tensammetric waves respectively; C_T^+ , C_T^- , corresponding tensammetric capacitance; R_T^+ , R_T^- , corresponding tensammetric leakage resistance; C_Δ , R_Δ , capacitance and leakage resistance of the electrode double layer during prevailing adsorption of dipoles; C_H , Helmholtz-Gouy capacitance of double layer.

the dipole and is equal and opposite to that of the same dipole when orientated through 180° (cf. Figs. 1 (a) and 1 (b)).

Figure 2 shows a typical tensammetric current-voltage curve. It can be seen that the curve consists essentially of two waves situated on either side of the electrocapillary zero point. In the following, these waves will be called the positive and the negative tensammetric wave respectively and their summit potentials will be distinguished by the symbols E_{s+} and E_{s-} . At high positive or high negative polarizations of the dropping mercury electrode, the tensammetric curve coincides with that of the supporting electrolyte. As is well known, maximum adsorption of the dipole occurs at the electrocapillary zero point and hence both the double-layer capacity and the A.C. current are appreciably lowered in this region (cf. Fig. 2).

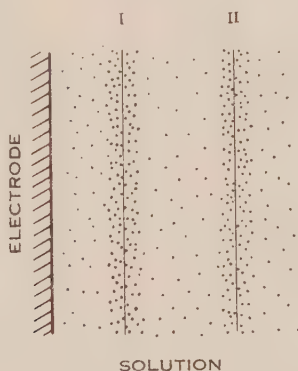


Fig. 3.—Distribution of "active charges" near the electrode-solution interface.

The reason for the large changes in differential current ($i_{D\sim}$) in the neighbourhood of E_{s+} and E_{s-} can be found in reorientation of the dipole molecules in the double layer with the change of the applied potential. These reorientation changes are discussed in detail in Section III (a).

Figure 2 shows also the various electrical equivalents of the electrode-double-layer system represented as lumped loss-free capacitors with leakage resistors in parallel. Following the nomenclature proposed by Grahame (1946), C_L represents the capacity and R_Δ the leakage resistance of an uncompressed layer of surface active molecules at the electrode-solution boundary.

III. THE SIGNIFICANCE OF THE TENSAMMETRIC POTENTIAL (E_s) AND OF THE TENSAMMETRIC CAPACITANCE ($C_{T\sim}$)

As shown in Section II, dipoles behave as "active charges" when in a non-uniform electric field like that encountered between the electrode surface and the bulk of the solution. Consider now these active charges to be distributed between regions I and II (Fig. 3) near the surface of the electrode. The picture is that of a variable condenser whose plates are situated at I and II and whose dielectric contains active charges.

It is the object of the following treatment to show that the capacitance of the variable condenser system becomes a maximum at the applied D.C. potential E_s .

Let the volume charge density of active charges be q_r^I (coulombs cm.³) in region I and q_r^{II} in region II and let their sum be equal to Q . The particles in the spaces I and II are in equilibrium

$$I \rightleftharpoons II. \quad \dots\dots\dots (1)$$

Let a^I and a^{II} denote the activities, γ^I and γ^{II} the activity coefficients, c^I and c^{II} the concentrations (moles/cm.³) of the active substance participating in the equilibrium process and contained in spaces I and II respectively. The dynamic capacitance is determined by the relation

$$C = \frac{dq_r}{d\psi}, \quad \dots\dots\dots (2)$$

ψ being the potential difference between the regions I and II. As we are only concerned with changes of ψ and not with its absolute value, we can write

$$C = \frac{dq_r}{dE}, \quad \dots\dots\dots (3)$$

where E is the applied potential between the electrode and the bulk of the solution.

If (1) is reversible then

$$Q_0 E = RT \ln K - RT \ln K', \quad \dots\dots\dots (4)$$

where K is the ratio of the activities of the charges in regions I and II at the potential $E = 0$, K' denotes the same ratio at any potential E , and Q_0 represents the charge associated with 1 gram mole of active substance. The relationship between Q_0 and Q is given by

$$Q_0 = \frac{Q}{c^I + c^{II}}, \quad \dots\dots\dots (5)$$

and

$$Q = |q_r^I| + |q_r^{II}|. \quad \dots\dots\dots (6)$$

If E_s represents that potential at which $a^I/a^{II} = 1$ the following relationship is obtained :

$$E = E_s - \frac{RT}{Q_0} \ln K'. \quad \dots\dots\dots (7)$$

Since

$$K' = \frac{\gamma^I c^I}{\gamma^{II} c^{II}} = \frac{q_r^I}{q_r^{II}}, \quad \dots\dots\dots (8)$$

and from (7)

$$K' = e^{-Q_0(E - E_s)/RT}, \quad \dots\dots\dots (9)$$

the dynamic capacitance of the condenser system follows as

$$C = \frac{dq_r}{dK'} \cdot \frac{dK'}{dE}, \quad \dots\dots\dots (10)$$

For small deviations of E from E_x , Q and Q_0 can be considered independent of E . Combining (6) and (8) we obtain

$$K' = \frac{Q}{q_v^{\text{II}}} - 1, \quad \dots \quad (11)$$

and

$$\frac{dK'}{dq_v^{\text{II}}} = -Q(q_v^{\text{II}})^{-2},$$

or

$$\frac{dq_v^{\text{II}}}{dK'} = -\frac{(q_v^{\text{II}})^2}{Q}. \quad \dots \quad (12)$$

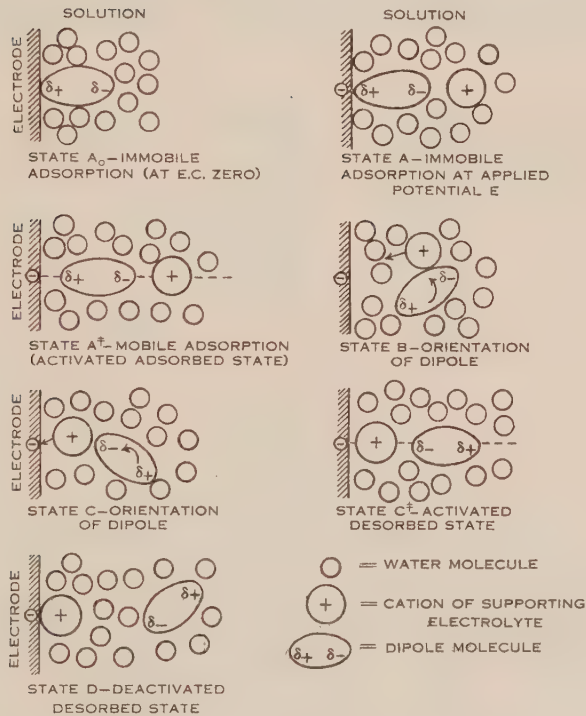


Fig. 4.—Schematic representation of the adsorption-desorption process.

Differentiating (9)

$$\frac{dK'}{dE} = -\frac{Q_0}{RT} e^{-Q_0(E-E_x)/RT}, \quad \dots \quad (13)$$

and combining (10), (12), and (13)

$$C = \frac{(q_v^{\text{II}})^2}{RT} \cdot \frac{Q_0}{Q} e^{-Q_0(E-E_x)/RT}. \quad \dots \quad (14)$$

Since

$$q_v^{\text{II}} = \frac{Q}{1 + e^{-\phi}}, \quad \dots \quad (15)$$

where

$$\varphi = \frac{Q_0(E - E_x)}{RT}, \dots\dots\dots (16)$$

it follows that

$$C = \frac{Q^2 \cdot e^{-\varphi}}{(c^I + c^{II})RT(1 + e^{-\varphi})^2} = \frac{QQ_0e^{-\varphi}}{RT(1 + e^{-\varphi})^2}, \dots\dots\dots (17)$$

C is therefore a maximum when $\varphi = 0$, that is, when $E = E_x$ and hence when $K' = 1$.

To a maximum capacitance corresponds a maximum A.C. current and it follows therefore from the tensammetric curve that E_x is identical with the summit potential E_s (Fig. 2). It can thus be concluded that at this point of the curve the concentrations (activities) of the active charges in regions I and II are equal.

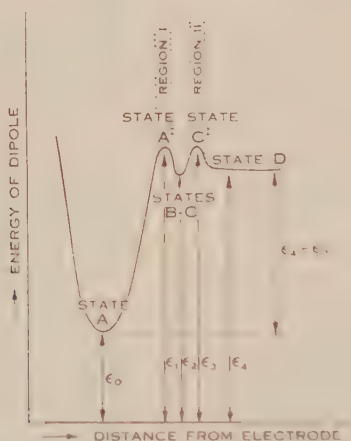


Fig. 5.—Energy states of dipole molecule during the adsorption-desorption process (after Fig. 4).

ϵ_0 , energy of dipole in adsorbed state A (immobile adsorption); ϵ_1 , energy of dipole in region I (state A^\ddagger activated adsorption); ϵ_2 , energy of dipole in states B and C (orientation); ϵ_3 , energy of dipole in region II (state C^\ddagger activated desorption); ϵ_4 , energy of dipole in state D (deactivated desorption).

(a) *The Tensammetric Potential (E_s)*

The equilibrium between the active charges in regions I and II is actually the equilibrium between adsorbed and desorbed dipoles. The mechanism of this adsorption-desorption process is shown in Figures 4 and 5 and discussed in the following.

At the electrocapillary zero point the greatest number of dipoles at the electrode surface is in a state of immobile adsorption (state A_0) (cf. Glasstone, Laidler, and Eyring 1941). When a potential is applied between the electrode and the bulk of the solution a double layer is formed consisting of ions and dipoles on the one hand and the electrode on the other. The arrangement is such that most of the dipoles find themselves between the electrode and the ions. At the same time the dipoles are distorted due to the presence of a

non-uniform electric field (state A). As the potential difference between solution and electrode is further increased, desorption begins. The majority of dipoles now find themselves in a state of "mobile adsorption" (cf. Glasstone, Laidler, and Eyring 1941) (state A^\ddagger), which is followed by reorientation (states B and C), by the activated desorbed state (state C^\ddagger), and finally by the deactivated desorbed state (state D).

The equilibrium between the active charges in regions I and II is actually the equilibrium between the *activated* adsorbed state A^\ddagger and the *activated* desorbed state C^\ddagger . The complete mechanism can be represented by



If reorientation is relatively fast compared with $A \rightleftharpoons A^\ddagger$ and $C^\ddagger \rightleftharpoons D$, then (18) can be written as



Since adsorption is the spontaneous process, a molecule in the immobile adsorbed state will possess minimum energy (ε_0 , Fig. 5). The desorption rate is determined by the number of molecules in state A per unit volume (c_A) and is given by (Glasstone, Laidler, and Eyring 1941):

$$v_{\text{Des}} = c_A \frac{kT}{h} \cdot \frac{F_{A^\ddagger}}{F_A} e^{-(\varepsilon_1 - \varepsilon_0)/kT}, \quad (20)$$

where k = Boltzmann's constant,

T = absolute temperature,

h = Planck's constant,

F_A = partition function of the molecules in state A ,

F_{A^\ddagger} = partition function of the molecules in state A^\ddagger ,

ε_0 = energy of molecule in state A ,

ε_1 = energy of molecule in state A^\ddagger .

The adsorption rate, on the other hand, will be determined by the number of molecules (c_D) in state D after the establishment of an equilibrium. Thus

$$v_{\text{Ads}} = c_D \frac{kT}{h} \cdot \frac{F_C^\ddagger}{F_D} e^{-(\varepsilon_3 - \varepsilon_4)/kT}, \quad (21)$$

in which F_D = partition function of the molecules in state D ,

F_C^\ddagger = partition function of the molecules in state C^\ddagger ,

ε_3 = energy of molecule in state C^\ddagger ,

ε_4 = energy of molecule in state D .

At E_s the adsorption and desorption rates are the same

$$v_{\text{Des}} = v_{\text{Ads}}, \quad (22)$$

and both the concentrations of active charges in regions I and II and their respective energy states are equal, so that

$$\varepsilon_1 = \varepsilon_2 \quad \text{and} \quad F_C^\ddagger = F_{A^\ddagger}, \quad (23)$$

and therefore

$$\frac{c_D}{c_A} = K_1 = \frac{F^A}{F^D} e^{(\varepsilon_A - \varepsilon_D)/kT}, \quad \dots\dots\dots (24)$$

or

$$K_1 = \frac{F^A}{F^D} e^{Q_0 E_s / RT}, \quad \dots\dots\dots (25)$$

where $Q_0 E_s$ represents the work done in desorbing 1 gram mole of dipole from state A to state D . It follows that

$$Q_0 E_s = RT \ln K_1 + RT \ln \frac{F^D}{F^A},$$

or

$$E_s = \frac{RT}{Q_0} (\ln K_1 + \ln K_2) = \frac{RT}{Q_0} \ln M, \quad \dots\dots\dots (26)$$

where

$$M = K_1 \cdot K_2 \text{ and } K_2 = F^D / F^A. \quad \dots\dots\dots (27)$$

E_s is expressed with respect to the electrocapillary zero point.

If Q_s represents the surface charge density (coulombs cm^2) of the electrode surface then the work done in desorbing the dipole from the electrode surface is given by

$$Q_s E_s = \frac{1}{2} (C_{H(E_s)} - C_{\Delta_0}) E_s^2,$$

that is

$$Q_s = \frac{1}{2} (C_{H(E_s)} - C_{\Delta_0}) E_s, \quad \dots\dots\dots (28)$$

where $C_{H(E_s)}$ refers to the capacitance of the rigid double layer at the potential E_s , and C_{Δ_0} represents the capacitance of the electrode-solution system at the electrocapillary zero, E^0 , respectively (both in F. cm^2). Regarding the adsorbed molecules as of roughly cylindrical shape, perpendicularly orientated to the surface of the electrode and having a charge separation of l cm., Q_s , the volume charge density at the electrode surface will be approximately equal to Q_s/l . Remembering that

$$Q_0 = \frac{Q}{(c^I + c^{II})},$$

we have

$$Q_0 = \frac{Q_s}{l(c^I + c^{II})} = \frac{1}{2} \frac{(C_{H(E_s)} - C_{\Delta_0})}{l(c^I + c^{II})} E_s, \text{ (coulombs/mole)} \quad \dots\dots\dots (29)$$

and since $c^I = c^{II}$ at E_s ,

$$Q_0 = \frac{(C_{H(E_s)} - C_{\Delta_0})}{4lc^{II}} E_s, \quad \dots\dots\dots (30)$$

The total number of active charges in regions I and II depends upon the applied potential and upon the nature and concentration of the ions at the interface. Hence the concentration of active charges ($c^I + c^{II}$) participating in the process at E_s , for a given supporting electrolyte, can be expressed as

$$c^I + c^{II} = Kf(E_s), \quad \dots\dots\dots (31)$$

where K is a dimensional constant. Considering the special case when $f(E_s) = E_s$ equation (31) becomes

$$e^I + e^{II} = KE_s \quad (31a)$$

Substituting equation (31a) in (29) we obtain

$$Q_0 = \frac{(C_{H(E_s)} - C_{\Delta_0})}{2lK} \quad (32)$$

The magnitude of Q_0 , the number of active charges associated with 1 gram mole of active dipole, depends upon the polarization of the dipole at the electrode interface so that

$$Q_0 = L[g(P_{Di}, P_{Oi})], \quad (32a)$$

where L is a dimensional constant and $g(P_{Di}, P_{Oi})$ is a function of the distortion polarization, P_{Di} and the orientation polarization P_{Oi} of the dipole at the interface.

The excess desorption work (Φ) done at the potential E_s over that done at the electrocapillary zero point, where $E = 0$, is connected with the respective bulk concentration of dipole, c_B (after Frumkin 1926) by

$$|\ln c_B|_0^{E_s} = \frac{\Phi S}{RT}, \quad (33)$$

where S is the surface area requirement of 1 gram molecular weight of dipole. Since Φ represents the additional work of desorption done by the electrical forces we can write

$$\Phi = \frac{1}{2}(C_{H(E_s)} - C_{\Delta_0})E_s^2, \quad (34)$$

and substituting in equation (33)

$$|\ln c_B|_0^{E_s} = \frac{S(C_{H(E_s)} - C_{\Delta_0})E_s^2}{2RT} \quad (35)$$

Combining (35) with (32) we obtain

$$Q_0 = \frac{RT |\ln c_B|_0^{E_s}}{lSK E_s^2} = \frac{RT |\ln c_B|_0^{E_s}}{VK E_s^2}, \quad (36)$$

where V is the volume occupied by 1 mole of dipole at the interface.

Substituting Q_0 in equation (26)

$$E_s = \frac{|\ln c_B|_0^{E_s}}{VK \ln M}, \quad (37)$$

or

$$E_s = \frac{\ln c_B - \text{const.}}{VK \ln M} \quad (38)$$

This equation shows that the tensammetric potential should be a logarithmic function of the bulk concentration of dipole, a result which was confirmed experimentally. Figure 6 shows that the relationship between E_s and $\log c_{\text{Bulk}}$ for pyridine, octyl alcohol, and cyclohexanol is linear as demanded by equation (38).

(b) *The Tensammetric Capacitance (C_T)*

Superimposing a small sinusoidal potential

$$\Delta V = \Delta V_0 \sin \omega\tau, \quad \dots\dots\dots (39)$$

onto the applied direct potential E_s , an A.C. current (i) is produced in accordance with

$$\Delta i = \Delta i_0 \cos (\omega\tau + \theta). \quad \dots\dots\dots (40)$$

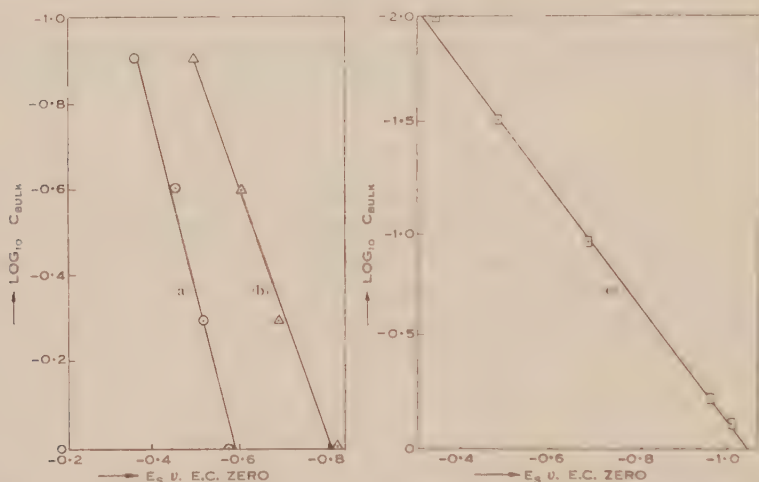


Fig. 6.—Relation between $\log c_{\text{Bulk}}$ and E_s for pyridine, octyl alcohol, and cyclohexanol.

(a) Octyl alcohol in 0.5N NaClO_4 ; T , 25 °C., $c_{\text{Bulk}} = 1$ for saturated octyl alcohol solution; (b) cyclohexanol in 0.5N NaClO_4 ; T , 25 °C., $c_{\text{Bulk}} = 1$ for saturated cyclohexanol solution; (c) pyridine in 0.1M KCl ; T , 25 °C., c_{Bulk} in molar concentrations.

Due to the applied A.C. field, the concentration of active charges at the electrode surface will vary sinusoidally, so that we can write

$$\Delta c = \Delta c_0 \cos (\omega\tau + \theta'), \quad \dots\dots\dots (41)$$

where θ and θ' represent the phase angles relative to the A.C. voltage of the current and of the concentration changes (Δc) respectively.

Displacement of dipoles from the electrode surface with increasing D.C. potential results in an excess concentration of dipoles near the electrode as compared with the bulk of the solution. Hence a concentration gradient is set up and the dipoles will diffuse from the electrode into the bulk. This diffusion process, which is slow compared with the rate of adsorption-desorption (cf. Frumkin and Melik-Gaikazyan 1951) is thus the rate determining step. By applying Fick's diffusion equation

$$\frac{\partial c}{\partial \tau} = D \frac{\partial^2 c}{\partial x^2}, \quad \dots\dots\dots (42)$$

and combining with equation (41) we obtain (after Randles 1947)

$$\frac{\partial (\Delta c)}{\partial x} \bigg|_{x=0} = -\frac{\Delta c}{2} \sqrt{\frac{\omega}{D}} \cos \left(\omega\tau + \theta' + \frac{\pi}{4} \right), \quad \dots\dots\dots (43)$$

where the distance x is reckoned from an imaginary boundary drawn across the region of active charges near the electrode and the bulk of the solution.

The alternating current is also given by $\Delta i = Q_0 A \cdot \text{flux}$, where

$$\text{flux} = D \frac{\partial(\Delta c)}{\partial x}, \dots\dots\dots (44)$$

A is the area of the electrode surface and D the diffusion coefficient of the dipole in the aqueous phase.

Combining (43) and (44) we obtain

$$\Delta i = Q_0 A \frac{\Delta c_0}{2} \sqrt{\omega D} \cos \left(\omega \tau + \theta' + \frac{\pi}{4} \right), \dots\dots\dots (45)$$

The amplitude of the A.C., Δi_0 , is therefore expressed as

$$\Delta i_0 = Q_0 A \frac{\Delta c_0}{2} \sqrt{\omega D}, \dots\dots\dots (46)$$

where

$$\theta = \theta' + \pi/4. \dots\dots\dots (47)$$

The concentration changes in virtue of the superimposed A.C. potential, ΔV , and at the applied D.C. potential, E_s , are given by

$$\Delta V_0 \cos \omega \tau = \frac{RT}{Q_0} \ln \frac{c - \Delta c_0 \cos(\omega \tau + \theta')}{c + \Delta c_0 \cos(\omega \tau + \theta')}, \dots\dots\dots (48)$$

where c refers to the concentrations of active charges in regions I and II which at E_s are equal, that is, $c^I = c^{II} = c$.

Rearranging,

$$\Delta c_0 \cos(\omega \tau + \theta') = -c \tanh \left(\frac{\psi}{2} \right), \dots\dots\dots (49)$$

where

$$\psi = \frac{Q_0 \Delta V_0 \cos \omega \tau}{RT}. \dots\dots\dots (50)$$

At $\tau = 0$

$$\Delta c_0 = \frac{c}{\cos \theta'} \cdot \tanh \left(\frac{\psi_0}{2} \right), \dots\dots\dots (51)$$

where

$$\psi_0 = \frac{Q_0 \Delta V_0}{RT}. \dots\dots\dots (52)$$

Applying a theoretical treatment similar to that used in deriving the A.C. active space and the dynamic capacitance of electrode reactions in alternating fields (cf. Breyer and Hacopian 1952*a*, 1952*b*) the tensammetric capacitance results as

$$C_T \sim \frac{\pi Q_0 b (mt)^{2/3} D^{1/2} c}{8 \sqrt{2} \Delta V_0 \omega^{1/2}} \left(\tanh \frac{\psi_0}{2} \right) (1 - \tan \theta'), \dots\dots (53)$$

where $b = 8.5 \times 10^{-1}$, involving the specific gravity of mercury and having the dimensions $M^{-2/3} L^2$, m is the mass of mercury (g./sec.), and t is the drop time of the capillary (in sec.).

Combining (53) with (31a) and (32a) we obtain

$$C_{T\sim} = \frac{\pi K b (m t)^{2/3} D^{1/2} L [g(P_{Di}, P_{Oi})] E_s}{8 \sqrt{2} \Delta V_0 \omega^{1/2}} \left(\tanh \frac{j\theta}{2} \right) [1 - \tan(\theta - \pi/4)]. \quad (54)$$

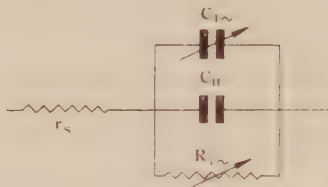


Fig. 7.—Equivalent circuit of the electrode system at E_s . r_s , ohmic series resistance of the cell; $C_{T\sim}$, tensammetric capacitance; C_H , Helmholtz-Gouy capacitance; $R_{T\sim}$, leakage resistance at the standard tensammetric potential.

The phase angle of the A.C., θ , is a measure of the reversibility of the adsorption-desorption process at the particular A.C. frequency (ω) and varies between $\pi/4$ and $\pi/2$ (47). For a perfectly reversible process, concentration

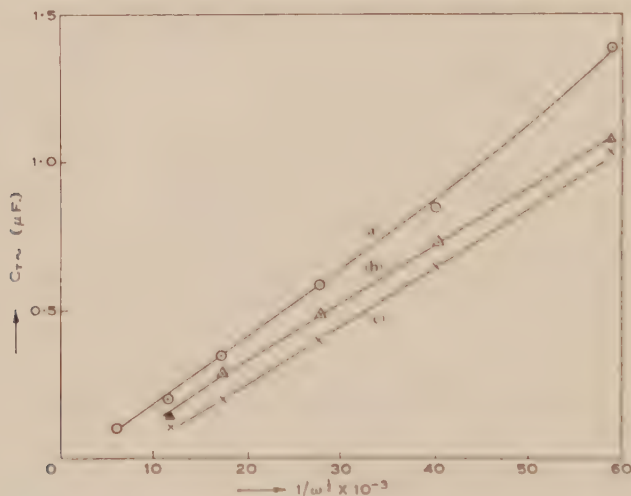


Fig. 8.—The linear relationship between the tensammetric capacitance ($C_{T\sim}$) and $1/\omega^{1/2}$ for pyridine, octyl alcohol, and cyclohexanol. A.C. 10 mV. r.m.s.; T , 20 °C.; capillary characteristics: m , 1.5 mg. sec.⁻¹, t , 3.5 sec.
(a) 0.1M pyridine in 0.1 KCl; (b) saturated octyl alcohol in 0.5N NaClO₄;
(c) saturated cyclohexanol in 0.5N NaClO₄.

changes and A.C. voltage are in phase so that $\theta = \pi/4$ and $\theta' = 0$. With decreasing reversibility the charge transfer will lag behind the superimposed A.C. voltage and $\theta = \pi/4$ whilst $\theta' = 0$. θ can be determined experimentally by the relation

$$\tan \theta = \omega C_{T\sim} R_{T\sim}. \quad (55)$$

As shown in Figure 7, C_{\sim} represents the total capacity of the network and is equal to the sum of the Helmholtz-Gouy capacity (C_H) and the tensammetric capacitance ($C_{T\sim}$). $R_{T\sim}$ is the corresponding leakage resistance. $C_{T\sim}$ reaches a maximum when $\theta = \pi/4$.

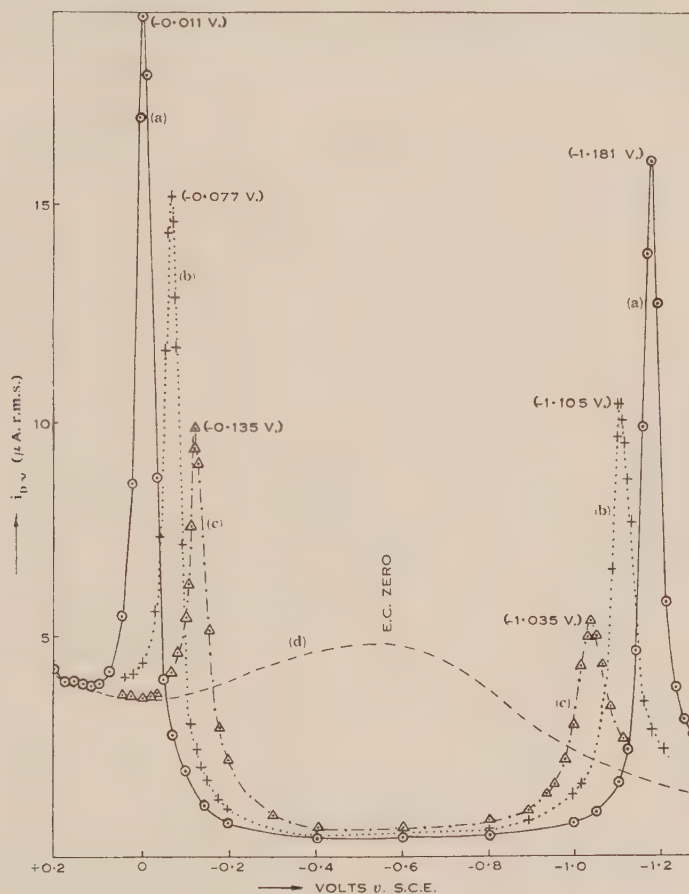


Fig. 9.—Tensammetric waves of octyl alcohol in 0.5N NaClO_4 . A.C. 15 mV. r.m.s.; T , 20 °C. (a) Saturated octyl alcohol in 0.5N NaClO_4 ; (b) half-saturated octyl alcohol in 0.5N NaClO_4 ; (c) quarter-saturated octyl alcohol in 0.5N NaClO_4 ; (d) supporting electrolyte alone (0.5N NaClO_4).

Equation (54) shows that the tensammetric capacitance is a measure of the distortion and orientation polarization of the molecule at the electrode interface. In addition, $C_{T\sim}$ appears to be linearly proportional to E_s and inversely proportional to the square root of the A.C. frequency. Since, however, P_{Di} , the distortion polarization of the dipole at the interface, is a function of the applied potential it is obvious that E_s and $C_{T\sim}$ will not in actual fact have a linear relationship. It follows also from equation (54) that no tensammetric wave can occur at the electrocapillary zero point as for $E_s \rightarrow 0$ also $C_{T\sim} \rightarrow 0$.

It should be stressed that the theoretical treatment given above applies for a given polarization only, that is, the values of $C_{T\sim}$ are different for one and the same substance at E_{s+} and E_{s-} . The reason is that K , L and the polarization function $g(P_{Di}, P_{Oj})$ ((31) and (32a)) assume different values depending, for

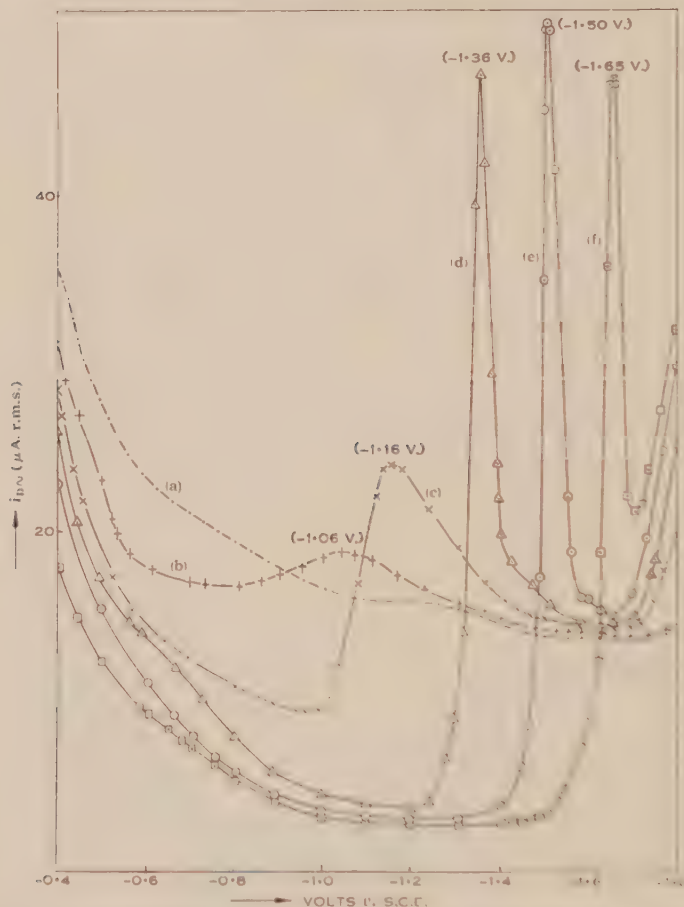


Fig. 10.—Tensammetric waves of pyridine in 0.1N KCl as supporting electrolyte. A.C. 45 mV. r.m.s.; T , 25 °C.

(a) 0.1N KCl; (b) 0.1N KCl+0.01M pyridine; (c) 0.1N KCl+0.05M pyridine; (d) 0.1N KCl+0.1M pyridine; (e) 0.1N KCl+0.2M pyridine; (f) 0.1N KCl+0.5M pyridine.

example, on the nature of the double layer which is different at positive and negative polarizations of the electrode. For the same reason also E_{s+} and E_{s-} are not, strictly speaking, equally removed from the electrocapillary zero point (E^0).

The relationship between ω and $C_{T\sim}$ for pyridine, octyl alcohol, and cyclohexanol is shown in Figure 8. A practically linear dependence as demanded by equation (54) was obtained.

IV. VARIATION OF TENSAMMETRIC SUMMIT POTENTIALS AND WAVE HEIGHTS WITH CONCENTRATION AND TEMPERATURE CHANGES

Saturated *cyclohexanol* solution in $N K_2SO_4$ (20 °C.) gave two well-defined tensammetric waves at $+0.045$ V. and -1.243 V. *v.* S.C.E. respectively. In half-saturated *cyclohexanol* solution the waves shifted to -0.015 and -1.159 V. *v.* S.C.E. respectively with corresponding reduction in the wave heights. As demanded by equation (38) both the positive and the negative waves shifted

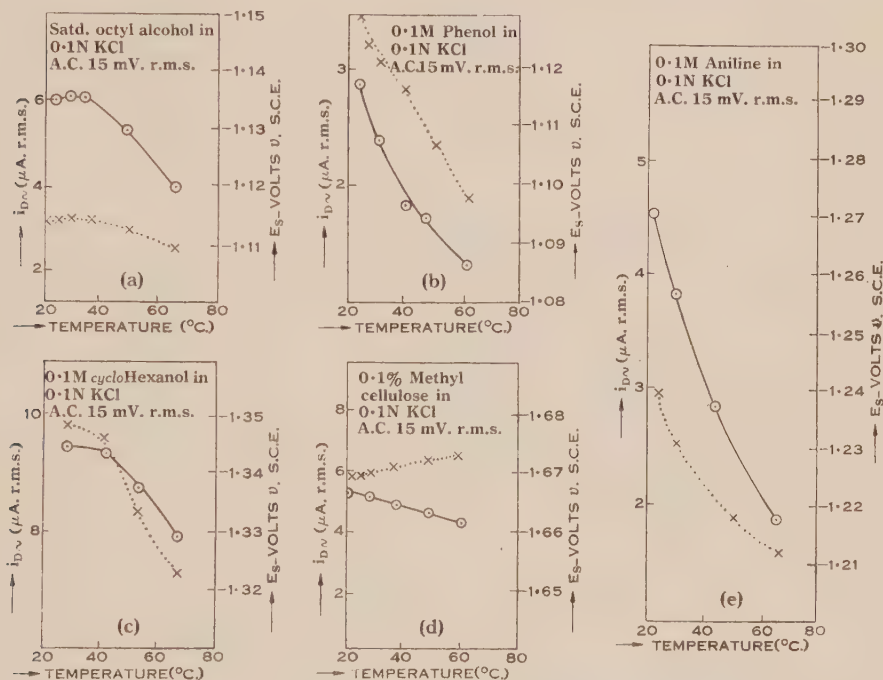


Fig. 11.—Variation of tensammetric wave height ($i_{D\sim}$) and negative tensammetric potential (E_{s-}) with temperature.

○—○ E_{s-} -temperature curves.

×.....× $i_{D\sim}$ -temperature curves.

towards the electrocapillary zero point on dilution. A similar behaviour was observed with saturated *cyclohexanol* solution in $0.5N NaClO_4$. The shifts of the positive and negative tensammetric potentials of octyl alcohol with dilution are shown in Figure 9.

Figure 10 shows the effect of dilution on the pyridine waves. A shift of as much as 0.6 V. was observed when the concentration of pyridine was changed from 0.5 to 0.01 molar (in 0.1 N KCl). Figure 10 also shows the corresponding increase of the base current in the region of the electrocapillary zero point with dilution of the pyridine, indicating that less pyridine is adsorbed at the interface as the dilution increases. The wave heights of pyridine were not greatly affected in the concentration range between 0.1 and 0.5 molar.

The influence of temperature changes on the wave heights of 2-methyl hexanol, phenol, *cyclohexanol*, methyl cellulose, and aniline are shown in Figures 11 (a), (b), (c), (d), and (e). It can be seen that generally both the wave heights and the E_c values decrease with increasing temperature. This behaviour

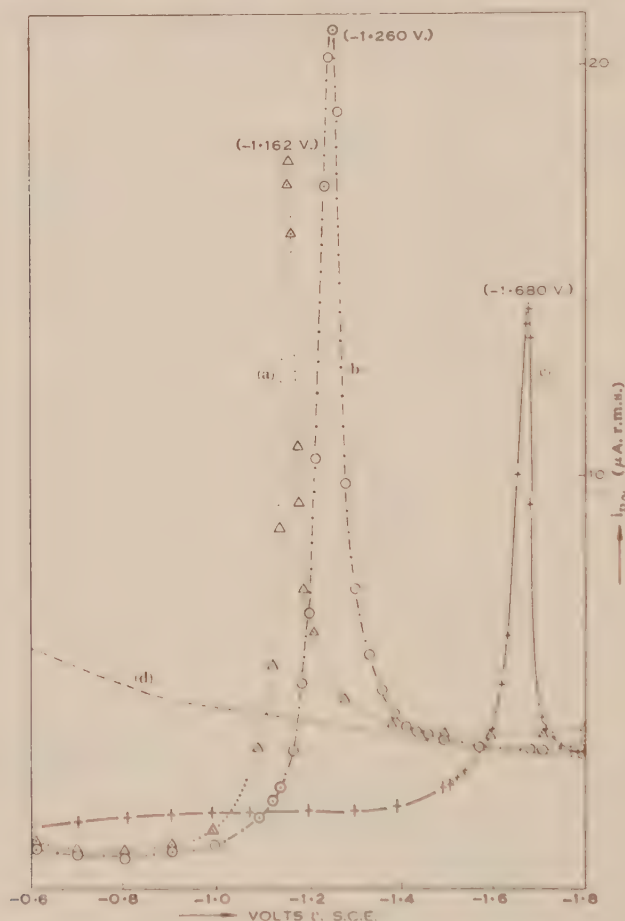


Fig. 12.—Influence of tensammetric waves on one another.
 (a) Saturated octyl alcohol in 0.1N KCl; (b) same as (a) + 0.1 M *cyclohexanol*; (c) same as (b) + 0.1 per cent. methyl cellulose;
 (d) 0.1N KCl (supporting electrolyte).

is best explained by remembering that both the kinetic energy and the solubility of the surface active substance increase with increasing temperature, facts which lead to easier desorption from the interface.

That solubility changes are at least partly responsible for the observed changes in tensammetric waves with change in temperature is shown by the behaviour of methyl cellulose: as is well known the solubility of this substance decreases with increasing temperature and it was found accordingly that in this case the wave height increased with a rise of temperature. The E_c values

changed but very little. The behaviour of weak acids (phenol, cresol, etc.) and weak bases (e.g. aniline, pyridine) also showed how the wave height varies with solubility changes : whereas the free acids and bases gave good tensammetric waves, no waves were obtained in media of such pH that the respective salts were formed. Due to the greater solubility of the salt compared with that of the free acid or base, practically complete desorption from the interface takes place.

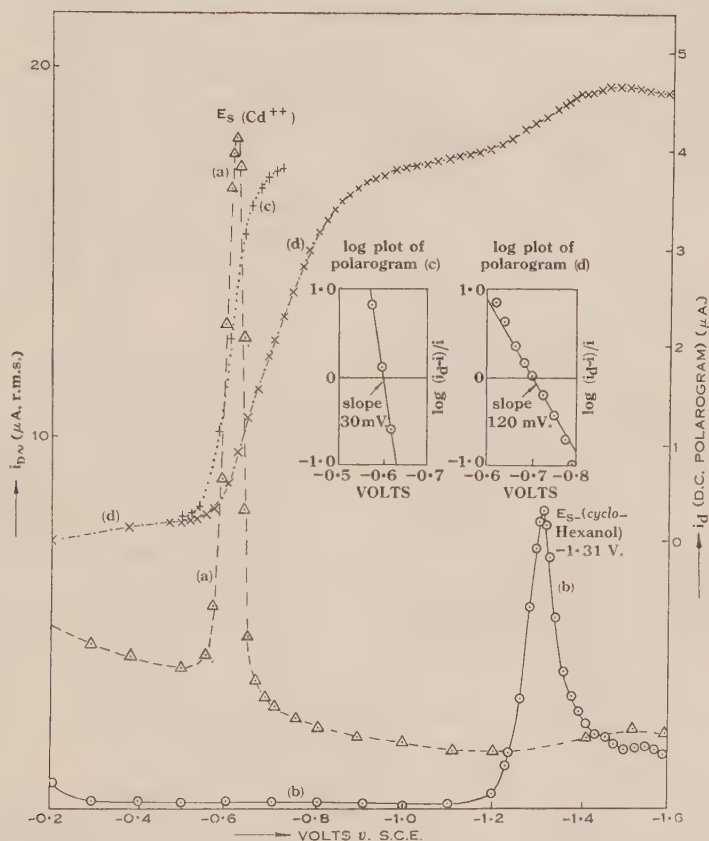


Fig. 13.—Effect of the presence of *cyclohexanol* on the A.C.-D.C. polarograms of Cd^{++} . The polarograms show that the discharge of Cd^{++} becomes irreversible in the presence of the adsorbed molecules of *cyclohexanol*.

(a) A.C. polarogram of 10^{-3}N Cd^{++} in 0.1N KCl (air-free) : A.C. 15 mV , r.m.s. ; T , 20°C .
 (b) A.C. polarogram of (a) + 0.1M *cyclohexanol* : only the tensammetric wave of *cyclohexanol* appears ; (c) D.C. polarogram of solution as under (a) ; (d) D.C. polarogram of solution as under (b).

V. INFLUENCE OF TENSAMMETRIC WAVES ON ONE ANOTHER AND ON CATHODIC A.C. AND D.C. POLAROGRAPHIC WAVES

Examining the region of negative polarizations, it was observed that when a mixture of two surface active substances was present in an indifferent supporting electrolyte, only one tensammetric wave was obtained corresponding to that

substance whose E_s occurred at a more negative potential. In other words, only that substance which was more strongly adsorbed at the mercury-solution interface produced a tensammetric wave. The phenomenon is illustrated in

TABLE 1

TENSAMMETRIC POTENTIALS (E_s) AND WAVE HEIGHTS (i_D) OF A NUMBER OF SURFACE ACTIVE SUBSTANCES IN VARIOUS SUPPORTING ELECTROLYTES (AIR-FREE)

Capillary characteristics: m , 1.5 mg. sec.⁻¹; t , 3.5 sec. (open circuit); A.C. 15 mV. r.m.s.; T , 25°C.

Compound	Concn.	Supporting Electrolyte	Tensammetric Potential		Max. Differential Current (i_D)	
			E_{s+} v. S.C.E. (V.)	E_{s-} v. S.C.E. (V.)	+ve Wave (μ A. r.m.s.)	-ve Wave (μ A. r.m.s.)
Octyl alcohol ..	Satd.	N Na_2SO_4	0.00	-1.140	17.1	4.2
" " ..	"	N NaNO_3	0.00	-1.140	30.5	9.2
" " ..	"	N KCl	—	-1.15	—	9.3
" " ..	"	N MgSO_4	0.00	-1.16	17.5	6.5
cycloHexanol ..	Satd.	N Na_2SO_4	+0.06	-1.30	34.0	10.8
" ..	"	N NaNO_3	-0.105	-1.18	14.0	8.2
" ..	"	N KCl	—	-1.28	—	14.0
" ..	"	N MgSO_4	+0.06	-1.35	34.6	9.1
n-Propyl alcohol	M	N Na_2SO_4	-0.015	-1.290	22.0	3.7
n-Butyl alcohol ..	0.1M	"	—	-1.16	—	3.2
Octyl alcohol ..	0.1M	0.1N KCl	—	-1.135	—	7.3
Benzyl alcohol ..	Satd.	N Na_2SO_4	+0.275	-1.450	24.5	13.5
Anisole ..	"	"	0.0	-1.140	5.5	6.5
Phenol ..	0.1M	"	—	-1.130	—	3.0
o-Cresol ..	Satd.	"	—	-1.270	—	10.5
Catechol ..	0.1M	"	—	-1.100	—	3.6
Resorcinol ..	"	"	—	-1.070	—	3.6
Hydroquinone ..	"	"	—	-1.060	—	3.6
Pyrogallol ..	M	"	-0.02	-1.290	<0.5	3.0
Benzoic acid ..	0.01M	"	—	-1.015	—	6.5
Gallic acid ..	Satd.	"	—	-1.080	—	<0.5
Aniline ..	0.1M	"	-0.03	-1.335	9.0	7.3
Pyridine ..	"	"	—	-1.383	—	32.5
Piperidine ..	Satd.	10N NaOH	—	-1.413	—	15.5
Methyl cellulose ..	0.1%	0.1N KCl	—	-1.665	—	6.0
" " ..	Satd.	"	—	-1.681	—	20.5

Figure 12. Curve (a) shows the negative tensammetric wave of saturated octyl alcohol in 0.1N KCl with the summit potential at -1.162 V. v. S.C.E. In the presence of 0.1M cyclohexanol, however, curve (a) disappears and only a single wave (b), corresponding to the tensammetric wave of cyclohexanol alone ($E_{s-} = -1.260$ V. v. S.C.E.), is obtained, probably because cyclohexanol is more

strongly adsorbed at the electrode surface than octyl alcohol. Adding now approximately 0.1 per cent. methyl cellulose to the mixture, the *cyclohexanol* in turn is displaced by the more strongly adsorbed methyl cellulose molecules as shown by the disappearance of the *cyclohexanol* wave and the appearance of the methyl cellulose wave only ($E_{s-} = -1.680$ V. *v.* S.C.E.).

In order to test the validity of the schematic representation of the various electrical equivalents of the electrode-double-layer system as shown in Figure 2, experiments were carried out studying the influence of tensammetric waves on A.C. and D.C. polarography. The results are illustrated in Figure 13. A.C. polarogram (a) was obtained by the reversible discharge of 10^{-3} N Cd^{++} in 0.1N KCl (air-free solution). Its corresponding D.C. polarogram is represented by curve (c). Plotting $\log(i_d - i)/i$ against E gave a straight line with a nearly theoretical slope of 30 mV./log unit. In the presence of *cyclohexanol*, however, the discharge of Cd^{++} was rendered irreversible as shown by the absence of the A.C. polarogram. Only the tensammetric wave of *cyclohexanol* ($E_{s-} = -1.31$ V.*v.* S.C.E.) appeared. The D.C. polarogram also revealed that the Cd^{++} reduction had become irreversible: the "log plot" of the Cd^{++} step was non-linear with an approximate slope of 120 mV. In contradistinction to Cd^{++} the discharge of Mn^{++} (-1.5 V. *v.* S.C.E.) was completely unaffected by the presence of *cyclohexanol*. The experiments clearly demonstrate that a reversible reduction process occurring within the region of dipole adsorption is rendered irreversible by the presence of a surface film at the electrode. In the case of Mn^{++} the reduction occurs at a potential at which the film is already desorbed (Fig. 2) and consequently both A.C. and D.C. polarograms remain unaffected. Thus in the presence of an adsorption layer the deposition of ions is impeded owing to the presence of a charged barrier at the electrode (see, for example, Heyrovsky, Sorm, and Forejt 1947; Loshkarev and Kryukova 1951; Maites and Maites 1951).

Tensammetric potentials and wave heights of a number of substances in various supporting electrolytes are given in Table 1.

VI. ACKNOWLEDGMENT

The authors are indebted to the Trustees of the Science and Industry Endowment Fund for a research grant.

VII. REFERENCES

- BREYER, B., and GUTMANN, F. (1946).—*Trans. Faraday Soc.* **42**: 645.
BREYER, B., and GUTMANN, F. (1947).—*Discuss. Faraday Soc.* No. 1 p. 19.
BREYER, B., GUTMANN, F., and HACOBIAN, S. (1950).—*Aust. J. Sci. Res. A* **3**: 558.
BREYER, B., and HACOBIAN, S. (1952a).—*Aust. J. Sci.* **14**: 118.
BREYER, B., and HACOBIAN, S. (1952b).—*Aust. J. Sci.* **14**: 153.
BUTLER, J. A. V. (1929).—*Proc. Roy. Soc. A* **122**: 399.
FRUMKIN, A. N. (1926).—*Z. Phys.* **35**: 792.
FRUMKIN, A. N., and GORODETSKAYA, A. (1928).—*C.R. Acad. Sci. U.R.S.S.* **18**: 639.
FRUMKIN, A. N., GORODETSKAYA, A., and CHUGUNOV, P. (1934).—*Acta Physiochim. U.R.S.S.* **1**: 12.
FRUMKIN, A. N., and MELIK-GAIKAZYAN, V. I. (1951).—*C.R. Acad. Sci. U.R.S.S.* **77**: 855.
(*Chem. Abstr.* **45**: 6909.)

- GLASSTONE, S., LAIDLER, K. J., and EYRING, H. (1941).—"Theory of Rate Processes." 1st Ed. p. 347 *et seq.* (McGraw-Hill Book Co. Inc.: New York and London.)
- GORODETSKAYA, A. (1940).—*Acta Physiochim. U.R.S.S.* **12**: 309.
- GOUY, G. (1903).—*Ann. Chim. Phys.* **29**: 145.
- GOUY, G. (1906a).—*Ann. Chim. Phys.* **8**: 291.
- GOUY, G. (1906b).—*Ann. Chim. Phys.* **9**: 75.
- GRAHAME, D. C. (1946).—*J. Amer. Chem. Soc.* **68**: 301.
- GRAHAME, D. C. (1947).—*Chem. Rev.* **41**: 441.
- HEYROVSKY, J. (1934).—*Actualités Sci. Industr. Paris* No. 90.
- HEYROVSKY, J., STORM, F., and FOREJT (1947).—*Coll. Trav. Chim. Tchécosl.* **12**: 11.
- ILKOVIČ, D. (1936).—*Coll. Trav. Chim. Tchécosl.* **8**: 13.
- KOLTHOFF, I. M., and LINGANE, J. J. (1941).—"Polarography." 1st Ed. pp. 16, 125. (Interscience Publishers Inc.: New York.)
- LOSHKAREV, M. A., and KRYUKOVA, A. (1951).—*Zhur. Anal. Khim.* **6**: 166. (*Chem. Abstr.* **45**: 8919.)
- MAITES, L., and MAITES, T. (1951).—*J. Amer. Chem. Soc.* **73**: 177.
- PROSKURNIN, M., and FRUMKIN, A. N. (1935).—*Trans. Faraday Soc.* **31**: 110.
- RANDLES, J. E. B. (1947).—*Discuss. Faraday Soc.* No. 1.

ISOTOPE EFFECT IN THE REACTION OF ^{14}C -FORMALDEHYDE WITH DIMEDONE

By A. M. DOWNES*

[Manuscript received February 28, 1952]

Summary

Formulae for the specific activity of a labelled product at any stage in a reaction which involves an "isotope effect" are discussed. The investigation of isotope effect by isolating a labelled product in consecutive fractions is illustrated by the reaction of ^{14}C -formaldehyde with dimedone. In this reaction the ^{14}C -formaldehyde reacts 0.927 ± 0.005 times as rapidly as the unlabelled compound at room temperature.

Care must be taken when using dimedone as a derivative for the accurate assay of isotopic formaldehyde if errors due to isotope effect are to be avoided.

I. INTRODUCTION

Values for the relative reaction rates of isotopic molecules in irreversible reactions may be determined by studying the changes in isotopic composition of the reactant and product at various stages of the reaction. Previous ways of doing this will be summarized below and a modified method will be introduced and exemplified.

Throughout the following discussion the simplest type of reaction will be considered, that in which only one of the reactants and one of the products are labelled. Such processes involve no overall isotope effect if the reaction is taken to completion. For more complicated reactions the situation may be different. For example, if one labelled reactant gives rise to two or more labelled products, the label may not be evenly distributed among the product molecules. Examples of such intramolecular isotope effects are the decarboxylations of carboxyl-labelled malonic acid (Bigeleisen and Friedman 1949; Yankwich and Calvin 1949; Lindsay, Bourns, and Thode 1951; Roe and Hellmann 1951) and oxalic acid (Lindsay, McElcheran, and Thode 1949). However, the equations given below apply to these cases also, provided the calculations are based on the mean specific activity of all the products or on the specific activity of one of them, corrected for intramolecular isotope effect. For convenience, the discussion will be in terms of "specific activities" defined as the ratio of the concentration of labelled to unlabelled molecules.

Let the specific rate constants for the reaction of the labelled and unlabelled reactant molecules be k^* and k . The ratio of these rate constants, k^*/k or ϵ , may be evaluated in several ways:

(a) The specific activity S_Y of the residual reactant at various stages of the reaction may be determined. The equation relating S_Y to the initial specific

* C.S.I.R.O. Tracer Elements Investigations, Melbourne.

activity S_0 and the fraction γ of total reaction has been given by Downes and Harris (1952) in a study of isotope effect in the Cannizzaro reaction of formaldehyde :

$$S_\gamma/S_0 = (1-\gamma)^\varepsilon. \quad (1)$$

This expression and those that follow are only valid for low isotopic concentrations, and are independent of the kinetic order of the reaction, since the labelled molecules are assumed to be present in tracer amounts in the derivation of the equation.

(b) The isotopic composition $S_{0\gamma}$ of the cumulative product up to the fraction γ of the complete reaction changes in the following way with γ :

$$S_{0\gamma}/S_{01} = [1 - (1-\gamma)^\varepsilon]/\gamma, \quad (2)$$

where S_{01} is the specific activity of the total product. This equation was derived by Bigeleisen (1949*b*) and, in a different form, by Stevens and Attree (1949). S_{01} must, of course, correspond with S_0 in (1) since there is no overall isotope effect for complete reaction.

When γ is small

$$S_{0\gamma}/S_{01} \doteq \varepsilon. \quad (3)$$

Stevens and Attree (1949) found ε for the hydrolysis of ^{14}C -carboxyl-labelled ethyl benzoate by applying equation (2). The decarboxylation of mesitoic acid (Bothner-By and Bigeleisen 1951) and of the trichloroacetate ion (Bigeleisen and Allen 1951) have been similarly studied using measurements of the natural ^{13}C abundance. In other investigations the first few per cent. of product have been isolated and ε obtained using the approximation (3) (Bigeleisen and Friedman 1949, 1950 ; Lindsay, McElcheran, and Thode 1949 ; Lindsay, Bourns, and Thode 1951).

(c) The product may be isolated in separate, consecutive fractions. This modification of (b) has not been used previously to find values of ε but has been employed to detect isotope effects qualitatively (see Section IV). The discussion and illustration of this method constitute the major portion of this paper.

II. DETERMINATION OF ε BY FRACTIONAL ISOLATION OF THE PRODUCT

Bigeleisen and Allen (1951) have discussed the difficulty of obtaining accurate values of ε from (2) using radioactive assay of ^{14}C . They claim that similar work with ^{13}C using mass-spectrometric analyses gives results one order of magnitude better. The maximum difference between $S_{0\gamma}$ and S_{01} occurs at the beginning of the reaction where $S_{0\gamma}/S_{01} = \varepsilon$ (see Fig. 1) and, since ε will be close to unity, the comparison of two specific activities of the same order of magnitude results in a large error in ε . However, the accuracy of the results may be increased by isolating the product as a series of consecutive fractions. An expression follows from (2) for the specific activity, symbolized as $S_{\gamma\gamma'}$, of the product formed between any two stages, γ_γ and $\gamma_{\gamma'}$, of the reaction :

$$R = \frac{S_{\gamma\gamma'}}{S_{01}} = \frac{(1-\gamma_\gamma)^\varepsilon - (1-\gamma_{\gamma'})^\varepsilon}{\gamma_{\gamma'} - \gamma_\gamma}. \quad (4)$$

Values of R calculated for $\varepsilon=0.9$ and for 0.1 intervals in γ are represented by the series of horizontal lines in Figure 1. It is obvious from the graph that much larger differences in specific activity than would be possible without fractional isolation are obtainable in this way. Theoretically, then, ε may be determined from the specific activity of each fraction using (4). In practice, however, the evaluation would be laborious since (4) cannot be solved for ε directly. Also the probable error in ε would be very large for the samples whose specific activities approximate S_{01} . The following modified procedure is more convenient and accurate.

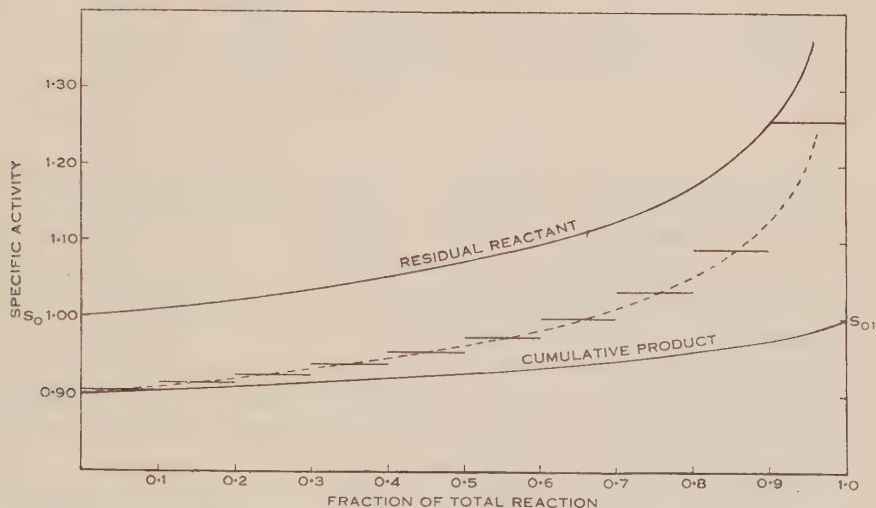


Fig. 1.—The changes in specific activity of a reactant and product as reaction proceeds, calculated for $\varepsilon=0.9$. The broken line shows the calculated specific activities of successive infinitesimal fractions of the product. The horizontal lines represent the specific activities of successive tenths of the product.

If the intervals between γ_x and γ_y are made infinitesimal (so that $\gamma_x \doteq \gamma_y = \gamma_m$) it can be shown that (4) becomes

$$R' = S_m/S_{01} = \varepsilon(1 - \gamma_m)^{\varepsilon-1}. \quad (5)$$

This expression, which gives the instantaneous specific activity of the product as it is formed, can be seen from (1) and (5) to be ε times the specific activity of the residual reactant at the same stage of the reaction. If the intervals between γ_x and γ_y are made small but finite and the mean of γ_x and γ_y taken as γ_m in (5), then (5) becomes an approximation of (4). Values of R and R' calculated by means of (4) and (5) for $\varepsilon=0.9$ are given in Table 1, which shows that agreement to within 0.5 per cent. up to 90 per cent. reaction is obtained, even for fractions as large as 0.2. Obviously the agreement becomes better as the fraction size is made smaller. The agreement is also seen in Figure 1 where the broken curve, representing R' as a function of γ , crosses the horizontal lines close to their midpoints, except when γ is greater than 0.9. For clarity all the above possible changes in specific activity have been plotted against γ in Figure 1. This illustrates the great magnification of the isotope effect which results by studying

the residual reactant as the reaction proceeds or by isolating the product fractionally. The latter method produces the largest possible differences in specific activity.

TABLE 1
COMPARISON OF RELATIVE SPECIFIC ACTIVITIES CALCULATED FROM (4) AND (5)

γ_1	γ_2	γ_m	R	R'	% Error
0.0	0.2	0.1	0.90974	0.90953	0.023
0.2	0.4	0.3	0.93303	0.93268	0.038
0.4	0.6	0.5	0.96532	0.96460	0.075
0.6	0.8	0.7	1.01730	1.01518	0.209
0.7	0.9	0.8	1.06250	1.05716	0.504
0.8	1.0	0.9	1.17462	1.13303	3.650

Since errors up to 0.5 per cent. are less than those of experimentally-determined specific activities, (5) may be used as a good approximation of (4) within the limits indicated. Hence, by plotting $\log S_{\gamma} / r. \log (1 - \gamma_m)$ for values of γ up to c. 0.9 a straight line with a slope equal to $(\varepsilon - 1)$ should result. Extrapolation of this line to $\gamma = 0$ will give the value of εS_{01} and by comparing this with the overall specific activity S_{01} determined in a separate experiment, another value of ε is obtainable. It will be noted that, in contrast to method (b) above, the determination of $(\varepsilon - 1)$ by this fractional isolation method is independent of the separate measurement of the overall specific activity: this may be an advantage in non-quantitative reactions.

When the reaction product is isolated as only two fractions symbolized as 0γ and $\gamma 1$, (4) reduces to (2) for the first fraction and to (1) for the second. Then

$$R'' = \frac{S_{0\gamma}}{S_{\gamma 1}} \frac{1 - (1 - \gamma)^2}{\gamma} : (1 - \gamma)^2 = 1,$$

whence

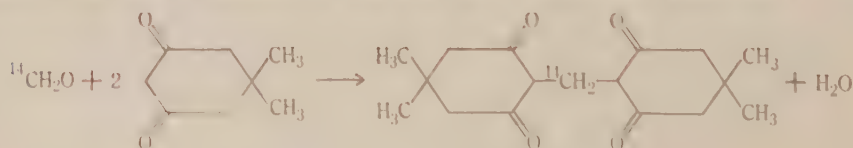
$$\varepsilon = 1 - \frac{\log (\gamma R'' + 1 - \gamma)}{\log (1 - \gamma)}, \quad \dots \dots \dots (6)$$

and this expression may also be used to evaluate ε .

The feasibility of these methods of determining ε depends on the ease of isolating a reaction product in clear-cut fractions. Where the product is a gas the techniques are comparatively simple. In other cases, where product isolation may be more difficult, the usefulness of this method is restricted.

III. RESULTS

Equations (5) and (6) have been tested by a study of the reaction of ^{14}C -formaldehyde with dimedone to form the solid bismethone derivative:



Using the conditions described by Yoe and Reid (1941) the quantitative isolation of formaldehyde bismethone in fractions is practicable. Two preliminary experiments, in each of which only two fractions were taken, indicated a substantial isotope effect in this reaction. From these experiments ϵ , calculated from (6), was shown to be $c. 0.93$.

Two more accurate experiments (Expts. 2 and 3) were then conducted, the bismethone being isolated as ten and five fractions respectively. The results are shown in Table 2. In another experiment several portions of labelled

TABLE 2
RADIOACTIVE ASSAY OF ^{14}C -FORMALDEHYDE BISMETHONE ISOLATED AS A
SERIES OF CONSECUTIVE FRACTIONS

Expt.	Sample	Weight (mg.)	Mean Fraction of Reaction γ_m	Specific Activity S_{xy} (c/m.)	$S_{xy}(\gamma_y - \gamma_x)$
2	1	91.7	0.055	1216	132.5
	2	94.7	0.165	1217	136.3
	3	88.3	0.273	1215	127.6
	4	90.1	0.380	1249	133.6
	5	89.9	0.486	1272	134.8
	6	81.8	0.588	1268	123.0
	7	85.8	0.687	1286	131.2
	8	83.4	0.787	1365	135.1
	9	78.8	0.884	1397	131.3
	10	58.7	0.966	1535	105.9
					Total 1291.3
3	1	91.5	0.109	1191	258.5
	2	93.3	0.328	1235	273.0
	3	90.1	0.546	1294	278.0
	4	88.3	0.758	1337	279.5
	5	57.9	0.931	1456	201.0
					Total 1290.0

formaldehyde from the same stock were precipitated quantitatively as the bismethone to obtain the overall specific activity S_{01} .* This was 1292 ± 6 counts per minute (c/m.), the mean of seven results (1290, 1300, 1267, 1300, 1308, 1278, 1301 respectively). Another method of obtaining S_{01} is to sum $[S_{xy}(\gamma_y - \gamma_x)]$ calculated for each sample from the fractional precipitations. This has been

* Since the absolute values are not needed, counts per minute are recorded as they are proportional to the specific activities of the "infinitely" thick tracer-labelled samples used.

done in Table 2, the sums being 1291 and 1290 c.m. for Expts. 2 and 3 respectively, in excellent agreement with the value determined previously.

For the first nine samples of Expt. 2 and the first four of Expt. 3, a plot of $\log S_{xy}$ v. $-\log(1-\gamma_m)$ is shown in Figure 2. The straight line of best fit through these points was determined by the method of least squares. The greatest scatter of these points corresponds to an error of c. 1.8 per cent. in S_{xy} . The slope of the line thus obtained, corresponding to ($\varepsilon = 1$), is 0.074 ± 0.006 . Hence $\varepsilon = 0.926 \pm 0.006$.

The value of the specific activity for $\gamma = 0$, determined from the graph by extrapolation, was 1199 ± 5 c.m. Hence, from (3), $\varepsilon = (1199 \pm 5) / (1292 \pm 6) = 0.928 \pm 0.006$.

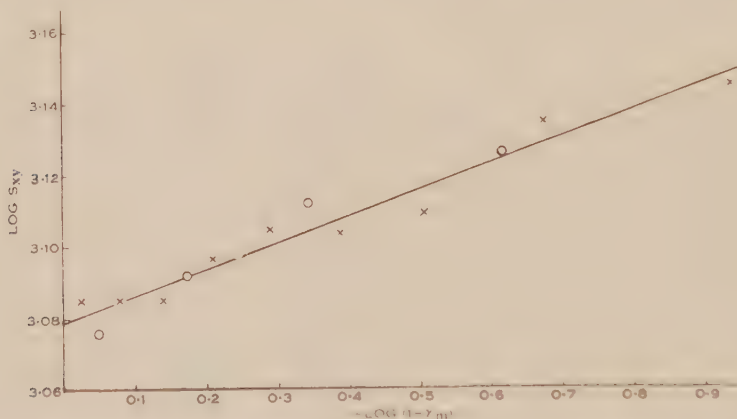


Fig. 2.—Isotope effect in the reaction of ^{14}C -formaldehyde with dimedone.

The values of ε determined by the two methods agree within the experimental error, and the mean may be taken as the best value, i.e. $\varepsilon = 0.927 \pm 0.005$.

The fact that an isotope effect was observed indicates that no rapid exchange occurs between formaldehyde and its dimedone derivative. However, the occurrence of a slow exchange would introduce errors. Check experiments were therefore conducted. They showed that no exchange was detectable over a period of 3 days under the reaction conditions used.

All the precipitations were conducted at room temperature (c. 23°C.) without any attempt at accurate temperature control. The results therefore probably include a small error due to the slight temperature dependence of ε (Bigeleisen 1949a; Lindsay, Bourns, and Thode 1951).

IV. DISCUSSION

The results obtained above show that the method of fractional isolation of the product can produce values of ε for ^{14}C reactions with good accuracy, the actual precision attainable depending on the number of samples taken.

Armstrong *et al.* (1950) have used a similar method to demonstrate an "isotope effect" in wet oxidations of ^{14}C -urea and its nitrate, although these authors were not interested in obtaining values of ε for the oxidations concerned.

They isolated the carbon dioxide produced in the oxidation as two consecutive fractions and showed the second fraction to be more active than the first. It is of interest to calculate ε from their results using (6). Good agreement in the values of ε (0.928, 0.923, 0.909; mean 0.920) for the three urea oxidations is shown, but the results for urea nitrate are not so consistent (0.974, 0.913). However, such a variation is within the experimental error for results based on two samples only in each oxidation.

The same technique has been applied by Evans and Huston (1951) to wet oxidations of ^{14}C -carboxyl-labelled acetic acid. They report an average value of 1.06 ± 0.02 for the ratios of specific activities of the second fractions of carbon dioxide to those of the first. It should be observed that their results are not necessarily due to an isotope effect at all but may simply be a manifestation of different rates of oxidation of the methyl and carboxyl carbons. This type of effect was reported by Armstrong *et al.* (1950) for the wet oxidation of ^{14}C -carbonyl-labelled xanthhydrol ureide where practically all the activity was obtained in the last 23 per cent. of the carbon dioxide evolved.

In view of the wide use of the dimedone derivative of ^{14}C -formaldehyde for radiochemical assay, the possible errors due to the isotope effect in its preparation should be stressed. Only by precipitating a sample of formaldehyde quantitatively can reliable results be obtained.

Under the experimental conditions used in the reaction with dimedone over 99 per cent. of the formaldehyde would exist as the hydrated monomer, methylene glycol (Auerbach, cited by Walker 1944, p. 31; Delépine 1897; Walker 1931). The observed isotope effect cannot, therefore, be due to depolymerization of a polyoxymethylene.

The kinetic studies of Wadano, Trogus, and Hess (1934) show that a state of equilibrium in aqueous formaldehyde solutions is reached in a few hours at room temperature. The added labelled formaldehyde had ample time to come to equilibrium before the reactions with dimedone were conducted. This is confirmed by the fact that no detectable difference in the magnitude of the isotope effect was observed over a period of about 6 weeks, using the same stock solution.

The only other possibility is that the effect occurs in the dehydration of methylene glycol to formaldehyde if this is necessary before the reaction with dimedone.

V. EXPERIMENTAL

The formaldehyde solution was made by heating paraformaldehyde (A.R., B.D.H.) in distilled water for several hours, cooling, filtering, and adding sufficient ^{14}C -formaldehyde* to give convenient counting rates. Iodometric analysis of the solution (Walker 1944) showed it to be 0.0952M. Dimedone (B.D.H.) was used without further purification.

The true specific activity of the formaldehyde was determined by quantitative precipitation of the bismethone (Yoe and Reid 1941), the compound being purified and counted as described below.

A typical fractional precipitation was carried out as follows: To a mixture of sodium acetate (15 ml.; 1N), hydrochloric acid (7.5 ml.; 1N), and labelled formaldehyde (30 ml.; 0.0952M),

* Obtained from the U.S. Atomic Energy Commission.

was added a saturated aqueous solution of dimedone (c. 15 ml.). The flask was stoppered and allowed to stand at least 24 hours. The formaldehyde bismethone was then filtered through a weighed sintered-glass funnel and washed with a little distilled water. The filtrate and washings were collected quantitatively, and to them were added more sodium acetate (5 ml. : 1N), hydrochloric acid (2.5 ml. : 1N), and dimedone solution (c. 15 ml.). The mixture was again allowed to stand 24 hours and the product filtered as before. This process was repeated until 10 fractions had been obtained. The pH of the filtrate from the final fraction was checked and found to be 4.40, close to the value 4.60 recommended by Yoe and Reid. The total weight of the samples in each experiment agreed within 1% with that calculated from the iodometric analysis. The error in specific activity due to the solubility of formaldehyde bismethone is thus negligible.

The funnels holding the product, after drying overnight in a vacuum desiccator containing sulphuric acid, were weighed. The samples were then transferred in ethanolic suspension to 30 ml. conical flasks, dissolved in a few ml. of hot aqueous ethanol, and precipitated by the rapid addition of cold water with swirling. The sudden precipitation produced small crystals in a state suitable for mounting prior to counting.

The samples were filtered and mounted using the method of Oddie and Lynn (1951), dried, and counted under a thin mica end-window Geiger-Müller tube with constant geometry. The samples, with an area of 2.61 cm.² and weighing c. 65 mg., were "infinitely" thick. An experimentally-determined self-absorption curve showed that 50 mg. was the minimum weight for infinite thickness. In a few cases where samples weighing less than 50 mg. were obtained self-absorption corrections were applied to give the infinite thickness value. Each sample was remounted twice, giving three values of the counting rate, the mean of which was taken for Table 2. Each count totalled at least 10,000, and the counting rates were corrected for counter sensitivity, resolving time of the tube, and background.

The radiochemical purity of the samples was established by the following facts:

- (i) Repeated crystallization from aqueous ethanol produced no change in specific activity.
- (ii) The m.p. (189–190 °C.) was sharp.
- (iii) The compound was completely soluble in dilute sodium hydroxide, indicating that no cyclization to the octahydroxanthene had occurred.

To demonstrate absence of exchange, ¹⁴C-formaldehyde bismethone (130 mg.) of known specific activity was added to a mixture of sodium acetate (10 ml. : 1N), hydrochloric acid (5 ml. : 1N), and inactive formaldehyde (10 ml. : 0.0952M), diluted to 50 ml. The mixture was shaken vigorously and allowed to stand 24 hours before filtering and re-counting. No change in specific activity was detectable. A similar experiment with inactive formaldehyde bismethone and active formaldehyde indicated no exchange over a period of 3 days.

VI. ACKNOWLEDGMENTS

The author is greatly indebted to Dr. G. M. Harris, Chemistry Department, University of Melbourne, and to Dr. T. H. Oddie, C.S.I.R.O. Tracer Elements Investigations, for many helpful discussions during the course of this work and in the preparation of the paper.

VII. REFERENCES

- ARMSTRONG, W. D., SINGER, L., ZBARSKY, S. H., and DUNSHEE, B. (1950).—*Science* **112**: 531.
BIGELEISEN, J. (1949a).—*J. Chem. Phys.* **17**: 425.
BIGELEISEN, J. (1949b).—*Science* **110**: 14.
BIGELEISEN, J., and ALLEN, T. L. (1951).—*J. Chem. Phys.* **19**: 760.
BIGELEISEN, J., and FRIEDMAN, L. (1949).—*J. Chem. Phys.* **17**: 998.
BIGELEISEN, J., and FRIEDMAN, L. (1950).—*J. Chem. Phys.* **18**: 1325.
BOTHNER-BY, A. A., and BIGELEISEN, J. (1951).—*J. Chem. Phys.* **19**: 755.
DELÉPINE, M. (1897).—*C.R. Acad. Sci. Paris* **124**: 1454.
DOWNES, A. M., and HARRIS, G. M. (1952).—*J. Chem. Phys.* **20**: 196.

- EVANS, E. A., and HUSTON, J. L. (1951).—*J. Chem. Phys.* **19**: 1214.
- LINDSAY, J. G., BOURNS, A. N., and THODE, H. G. (1951).—*Canad. J. Chem.* **29**: 192.
- LINDSAY, J. G., McELCHERAN, D. E., and THODE, H. G. (1949).—*J. Chem. Phys.* **17**: 589.
- ODDIE, T. H., and LYNN, K. R. (1951).—Commonw. X-Ray Radium Lab. Tracer Elements Rep. No. 1.
- ROE, A., and HELLMANN, M. (1951).—*J. Chem. Phys.* **19**: 660.
- STEVENS, W. H., and ATTREE, R. W. (1949).—*Canad. J. Res.* **27B**: 807.
- WADANO, M., TROGUS, C., and HESS, K. (1934).—*Ber. deutsch. chem. Ges.* **67**: 174.
- WALKER, J. F. (1931).—*J. Phys. Chem.* **35**: 1104.
- WALKER, J. F. (1944).—"Formaldehyde." p. 259. (Reinhold Publishing Corp.: New York.)
- YANKWICH, P. E., and CALVIN, M. (1949).—*J. Chem. Phys.* **17**: 109.
- YOE, J. H., and REID, L. C. (1941).—*Industr. Engng. Chem. (Anal. Ed.)* **13**: 238.

LIQUID-VAPOUR EQUILIBRIA

III. THE SYSTEMS BENZENE-*n*-HEPTANE, *n*-HEXANE-CHLOROBENZENE, AND *cyclo*HEXANE-NITROBENZENE

By I. BROWN*

[Manuscript received March 19, 1952]

Summary

Liquid-vapour equilibrium compositions and pressures and relevant thermodynamic functions are given for the systems benzene-*n*-heptane at 80 °C., *n*-hexane-chlorobenzene at 65 °C., and *cyclo*hexane-nitrobenzene at 80 °C. A comparison with values obtained previously for the first of these systems is used to demonstrate the improved performance of a new equilibrium still which is described. It is shown that consistent results can be obtained with the new still for miscible systems having a relative volatility in the range 1.0 to 14. Equilibrium data are also given for the system *n*-hexane-chlorobenzene at 760 mm. Hg to permit comparison with the performance of other equilibrium stills.

I. INTRODUCTION

The three binary systems studied were chosen to give a range of relative volatility from 1.0 to over 100, to determine the effect of relative volatility on the performance of the new equilibrium still. The need for a still that would give satisfactory performance with mixtures of high relative volatility was shown by the results of Brown and Ewald (1951) for the system benzene-*n*-heptane where inconsistency was found when the relative volatility was greater than 2.0.

II. APPARATUS

The equilibrium still of Brown and Ewald (1950), which will be referred to as still No. 1, suffered from two defects when used with the system benzene-*n*-heptane at low benzene concentrations. Errors in vapour concentration determinations were introduced by evaporation of small amounts of condensate from the vapour trap after the still was vented to atmosphere for sampling. Uneven boiling caused occasional hydraulic surges which led to temperature fluctuations and to difficulty in temperature measurement. With some systems of high relative volatility (e.g. acetone-water), it was also found that flash boiling occurred where the condensed vapour mixed with the hot liquid stream returning to the kettle.

These difficulties have been overcome by three major alterations to the still. The design of the new still, referred to as still No. 2, is shown in Figure 1.

The new type of sample trap has a small vapour hold-up, a small area of liquid free surface, and a magnetically operated ball valve to isolate the contents of the traps while the liquid and vapour streams are still circulating. This type

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

of trap also prevents contamination of the samples by liquid draining down from the walls after the still circulation is stopped, and it allows the contents of the traps to be cooled before sampling.

To promote even boiling, finely powdered Pyrex glass has been sintered on to the walls of the kettle, the taper to the Cottrell pump tube made more gradual, and the operating liquid level raised.

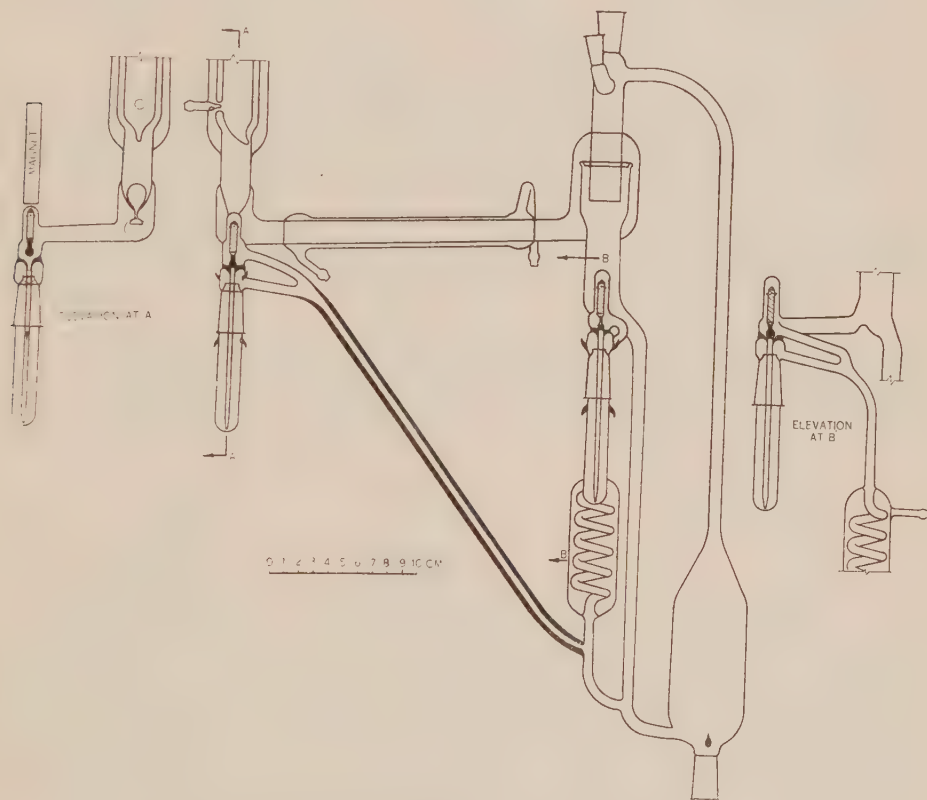


Fig. 1.—Equilibrium still No. 2.

Flash-boiling has been eliminated by the addition of a spiral heat-exchanger which cools part of the liquid return before it mixes with the condensed vapour that flows back to the kettle. A horizontal condenser aids cooling of the vapour sample and reduces the hold-up of liquid and vapour in the vertical condenser. A stoppered opening has also been provided for filling. The operating charge of the still is approximately 200 ml. of which 12 ml. is in each trap.

Operation of the still was commenced at atmospheric pressure and, when the traps filled, the pressure was gradually reduced to the required value or to that corresponding to the required temperature. The still was then allowed to function until a steady state was reached (1.5 to 4 hours); small adjustments of pressure were made if required. The valves on the traps were then closed,

the still vented to atmosphere through a drying tube, and the contents of the traps cooled and sampled for analysis. During the run the temperature and pressure were measured using the methods previously described.

III. PURIFICATION AND PROPERTIES OF COMPONENTS

(a) Purification of Materials

(i) *Benzene and n-Heptane*.—Details of the purification of these liquids and the physical properties of the products have already been given by Brown and Ewald (1951). The same samples were used in this work.

(ii) *n-Hexane*.—Spectroscopic grade *n*-hexane was fractionated at a reflux ratio of 75 to 1 through a 1.5 m., 25 mm. diameter column packed with 1.6 mm. Dixon packing. Samples of 10 ml. were collected and those with a boiling point of 68.72 to 68.75 °C. at 760 mm. Hg (uncorr.) and a refractive index n_D^{25} 1.3720 to 1.3723 were retained. This material was twice shaken for 5 hours at 45 °C. with redistilled chlorosulphonic acid to remove methylcyclopentane, as proposed by Shepard and Henne (1930), and the product, after separation, was washed with concentrated sulphuric acid, water, 10 per cent. sodium hydroxide, and water, and finally dried over anhydrous calcium chloride. The product was then refractionated through the same column at a reflux ratio of about 100 to 1; head and tail fractions were rejected. The physical properties of the product were as follows:

Property				This Sample	Rossini <i>et al.</i> (1947)
B.p./760 (°C.)	68.67	68.742
d_4^{25}	0.65480	0.65482
n_D^{25}	1.3722	1.37226

(iii) *Chlorobenzene*.—Reagent grade chlorobenzene was fractionated at a reflux ratio of about 30 to 1 through a 50 cm., 25 mm. diameter column packed with 3 mm. glass Fenske helices, and head and tail fractions were rejected. The product was shaken twice with 10 per cent. aqueous sodium hydroxide, washed, and dried over anhydrous calcium chloride, and refractionated through the same column at a reflux ratio of about 10 to 1. The physical properties of the product are given below and compared with values by various authors reported by Timmermans (1950).

Property				This Sample	Timmermans (1950)
B.p./760 (°C.)	131.72	131.687 to 132.00
d_4^{25}	1.10112	1.10110
n_D^{25}	1.5219	1.52190

(iv) *cycloHexane*.—Details of the purification and physical properties of this liquid have been described previously by Brown and Ewald (1950). The refractive index of the purified liquid was n_D^{25} 1.4233, as compared with the value of 1.42354 given by Rossini *et al.* (1947).

(v) *Nitrobenzene*.—Reagent grade nitrobenzene was shaken with three lots of 1N aqueous sodium carbonate, with 1N sulphuric acid, with water, and then dried over anhydrous calcium chloride. It was fractionated at 3 mm. Hg

at a reflux ratio of about 30 to 1 through the column used for the chlorobenzene; head and tail fractions were rejected. The physical properties of the product were as follows:

Property	This Sample	Timmermans (1950)
F.p. (°C.)	+5.69	5.67 to 5.77
B.p./760 (°C.).. ..	210.66	210.85
d_4^{25}	1.19833	1.1983 to 1.1986
n_D^{25}	1.5499	1.55006

(b) Vapour Pressures of Components

The vapour pressures of the components were measured in the equilibrium still using the same temperature and pressure measuring equipment as described previously. The results are shown in Table 1. The Antoine equation*

$$\log P = A - B/(C + t), \quad \dots\dots\dots (1)$$

was fitted to the vapour pressure data for each component by the method of Willingham *et al.* (1945) and the values of the constants found are given in Table 2.

TABLE 1
MEASURED VAPOUR PRESSURES OF PURIFIED COMPONENTS (STANDARD MM. Hg)

Benzene		<i>n</i> -Hexane		Chlorobenzene		Nitrobenzene	
<i>t</i> (°C.)	<i>P</i>	<i>t</i> (°C.)	<i>P</i>	<i>t</i> (°C.)	<i>P</i>	<i>t</i> (°C.)	<i>P</i>
40.00	182.93	36.44	243.09	62.04	72.43	134.10	82.23
50.00	271.52	41.05	291.20	62.06	72.48	139.75	100.30
60.00	391.74	44.45	331.27	66.38	86.69	145.17	120.64
70.00	551.25	47.32	368.46	74.13	117.91	149.73	140.20
80.00	758.44	48.98	391.25	81.60	156.48	154.61	164.01
80.07	760.00	50.60	414.71	89.06	204.36	159.77	192.80
<i>n</i> -Heptane		53.04	451.99	94.04	242.84	164.45	222.33
40.00	92.51	55.53	493.24	94.47	246.45	168.72	252.39
50.00	141.64	55.69	495.74	98.79	284.73	172.96	285.04
60.00	210.32	57.79	532.74	103.18	328.29	178.48	333.44
70.00	303.79	60.21	577.82	106.25	361.75	182.07	367.64
80.00	428.20	60.58	585.04	110.35	410.99	185.70	405.18
90.00	590.14	62.33	619.81	114.48	465.85	188.90	441.33
91.00	608.55	64.24	659.78	115.89	485.69	192.98	490.74
98.40	760.00	65.44	685.70	117.94	515.85	196.63	538.87
<i>cyclo</i> Hexane		65.76	692.93	121.10	565.34	200.41	592.53
30.00	121.91	67.55	733.75	123.39	606.63	203.88	645.62
40.00	184.75	67.67	736.17	126.18	652.69	206.62	690.23
50.00	272.02	68.70	761.15	128.45	695.11	209.49	739.54
60.00	389.60			130.37	732.69	210.626	759.98
70.00	544.23			131.40	753.43	210.629	760.04
80.72	760.00			131.70	759.54		
				131.73	760.28		

* A table of nomenclature is included at the end of this paper.

The standard deviations in fitting the vapour pressure data for these substances were less than the estimated experimental errors due to errors in the measurement of the temperature and pressure, which were $\pm 0.01^\circ\text{C}$. and ± 0.05 mm. Hg respectively.

TABLE 2
VALUES OF CONSTANTS IN ANTOINE EQUATION

Substance	<i>A</i>	<i>B</i>	<i>C</i>
Benzene	6.91279	1214.57	221.166
<i>n</i> -Heptane	6.88686	1258.27	215.701
<i>n</i> -Hexane	6.84625	1151.74	221.777
Chlorobenzene ..	6.99893	1444.75	219.113
<i>cyclo</i> Hexane ..	6.85366	1208.47	223.465
Nitrobenzene ..	7.08937	1726.66	199.613

(c) *Virial Coefficients and Molar Volumes*

The second virial coefficients in the equations of state for the vapours of the components were estimated from critical data using the Berthelot equation. The critical data for nitrobenzene were calculated by the method of Meissner (1949), which gave T_c 716°K. and P_c 47 atm. The virial coefficient values for benzene were those estimated by Tompa (1948). The values used are shown in Table 3.

The liquid molar volumes used in calculating the thermodynamic functions were obtained by graphical interpolation of published density data given by Timmermans (1950), I.C.T. (1928), and Egloff (1939). The values used are also given in Table 3.

TABLE 3
LIQUID MOLAR VOLUMES AND SECOND VIRIAL COEFFICIENTS (LITRES)

Substance	<i>t</i> (°C.)	<i>V</i>	β
Benzene	80	0.096	-0.980
<i>n</i> -Heptane	80	0.159	-1.514
<i>n</i> -Hexane	65	0.140	-1.227
Chlorobenzene ..	65	0.106	-1.633
<i>cyclo</i> Hexane ..	80	0.117	-1.085
Nitrobenzene ..	80	0.108	-2.089

IV. ANALYTICAL METHODS

The liquid and vapour samples from the system benzene-*n*-heptane were analysed by density measurements as described previously. The other systems were analysed by refractive index measurements. Mixtures were made by weighing the components, and compositions calculated allowing for air buoyancy

and for the weight of air-vapour mixture of the first component displaced on adding the second. The composition *versus* refractive index data that were used for the preparation of large-scale graphs are given in Table 4.

TABLE 4
COMPOSITION VERSUS REFRACTIVE INDEX

<i>n</i> -Hexane-Chlorobenzene		<i>cyclo</i> Hexane-Nitrobenzene	
x_1	n_D^{25}	x_1	n_D^{25}
0.0000	1.5219	0.0000	1.5499
0.1059	1.5031	0.1139	1.5342
0.1987	1.4872	0.2186	1.5202
0.2841	1.4733	0.2554	1.5155
0.3941	1.4557	0.3138	1.5076
0.4675	1.4445	0.4100	1.4951
0.6174	1.4222	0.5046	1.4828
0.7080	1.4098	0.6020	1.4704
0.8598	1.3894	0.7036	1.4580
0.9284	1.3809	0.7933	1.4467
1.0000	1.3722	0.8843	1.4362
		1.0000	1.4233

V. LIQUID-VAPOUR EQUILIBRIUM DATA

The equilibrium data were measured by the method described in Section II. The excess chemical potentials, the excess free energy of mixing at constant pressure, and the relative volatility were calculated from the measured liquid and vapour concentrations, pressures, and temperatures using equations (2) to (5).

$$\mu_1^E = RT \ln \gamma_1 = RT \ln (Py_1/P_1x_1) + f(V_1 - \beta_1)(P_1 - P), \quad \dots (2)$$

$$\mu_2^E = RT \ln \gamma_2 = RT \ln (Py_2/P_2x_2) + f(V_2 - \beta_2)(P_2 - P), \quad \dots (3)$$

$$G_{xp}^E = x_1\mu_1^E + x_2\mu_2^E, \quad \dots (4)$$

$$\alpha = y_1x_2/y_2x_1, \quad \dots (5)$$

The vapour pressures of the pure components were calculated from equation (1) using the values of the constants given in Table 2.

The values of the excess free energy of mixing at various concentrations were fitted by the method of least squares to equation (6) or the equivalent equation with b and c set equal to zero. Using the values of the constants so obtained, fitted values of the excess chemical potentials were calculated from equations (7) and (8).

$$G_{xp}^E = x_1x_2[a + b(x_1 - x_2) + c(x_1 - x_2)^2], \quad \dots (6)$$

$$\mu_1^E = x_2^2[a - b + c + (4b - 8c)x_1 + 12cx_1^2], \quad \dots (7)$$

$$\mu_2^E = x_1^2[a - 3b + 5c + (4b - 16c)x_1 + 12cx_1^2], \quad \dots (8)$$

(a) *The System Benzene-n-Heptane at 80 °C.*

The results obtained with this system using the new still are given in Table 5. The first three columns show the liquid and vapour concentrations and the pressure; the next four show the excess chemical potentials, the excess free energy of mixing, and the relative volatility. The last column shows the

TABLE 5
BENZENE-*n*-HEPTANE AT 80 °C. RESULTS WITH NEW STILL (No. 2)

x_1	y_1	P (mm. Hg)	μ_1^E (cal./mol.)	μ_2^E (cal./mol.)	G_{xp}^E (cal./mol.)	α	ΔG_{xp}^E (cal./mol.)
0.0464	0.0988	454.62	181.7	0.9	9.2	2.253	+0.1
0.0861	0.1729	476.25	172.1	2.0	16.6	2.219	+0.2
0.2004	0.3473	534.38	147.9	7.3	35.4	2.123	+0.2
0.2792	0.4412	569.49	126.6	13.8	45.3	2.038	-0.6
0.3842	0.5464	613.53	103.3	27.9	56.9	1.931	+0.1
0.4857	0.6304	650.16	78.6	49.4	63.6	1.806	+0.5
0.5824	0.7009	679.74	55.8	76.7	64.5	1.680	-0.0
0.6904	0.7759	708.78	36.1	112.0	59.6	1.553	-0.4
0.7842	0.8384	729.77	20.9	154.9	49.8	1.428	-0.1
0.8972	0.9149	748.46	4.8	241.9	29.2	1.232	+0.3

difference between G_{xp}^E calculated from equations (2) to (4) and that from equation (6). The values of the constants (cal. mol.) found in fitting the data to equation (6) were $a = +254.3$, $b = +64.3$, $c = +12.3$.

A comparison is given in Table 6 of the fitted values of G_{xp}^E calculated from equation (6) using these values of the constants, with the corresponding data

TABLE 6
BENZENE-*n*-HEPTANE AT 80 °C. COMPARISON OF G_{xp}^E (CAL. MOL.)

x_1	G_{xp}^E (Still No. 1)	G_{xp}^E (Still No. 2)	Difference (No. 1—No. 2)
0.1	21.9	19.0	2.9
0.2	38.9	35.2	3.7
0.3	51.7	48.4	3.3
0.4	60.4	58.1	2.3
0.5	65.0	63.6	1.4
0.6	65.1	64.2	0.9
0.7	59.9	59.2	0.7
0.8	48.3	47.6	0.7
0.9	28.9	28.2	0.7

obtained previously using still No. 1. The estimated error in G_{xp}^E was 2.0 cal. mol. and the standard deviation in fitting the results was 0.5 cal. mol. (No. 1) and 0.4 cal. mol. (No. 2). This comparison shows a significant difference between the two sets of results at concentrations of benzene below a mole

fraction of 0.4. This is the composition range where the relative volatility was highest for this system.

The thermodynamic consistency of the data obtained with the two stills was compared by considering the differences between the experimental values of the excess chemical potentials from equations (2) and (3) and the fitted values from equations (7) and (8). These differences $\Delta\mu^E$ are shown in Table 7. Before the significance of these differences could be determined it was necessary to evaluate the errors in the values of the excess chemical potentials due to errors in the measured quantities from which they were calculated. For temperature, pressure, and composition the errors were $\pm 0.01^\circ\text{C}$., ± 0.05 mm. Hg, and ± 0.00025 mol. fraction respectively. The corresponding errors, $\delta\mu^E$ in the excess chemical potentials also appear in Table 7.

TABLE 7
BENZENE-*n*-HEPTANE AT 80°C . ERRORS IN EXCESS CHEMICAL POTENTIALS (CAL./MOL.)

x_1	$\Delta\mu_1^E$ (No. 1)	$\Delta\mu_1^E$ (No. 2)	$\delta\mu_1^E$	$\Delta\mu_2^E$ (No. 1)	$\Delta\mu_2^E$ (No. 2)	$\delta\mu_2^E$
0.05 ,	-98	-9.0	3.8	+4	+0.6	1.2
0.1	-79	-9.0	3.2	+9	+1.0	1.4
0.5	-10	-4.5	1.7	+10	+5.0	1.9
0.9	-1	-0.1	1.0	+7	+4.0	3.4

These results show that the values obtained using still No. 2 are considerably more consistent than those obtained using still No. 1. The consistency of the results was also tested by the method of Redlich and Kister (1948); the experimental values of $(\mu_1^E - \mu_2^E)$ were plotted against x_1 and the areas above and below datum were measured. The ratio of these areas was 1.83 for still No. 1, while for the new still it was 1.09.

(b) *The System n-Hexane-Chlorobenzene at 65°C .*

The results obtained with this system are shown in Table 8. The accuracy of these excess free energy data did not justify the use of the three-constant equation (6). The data were fitted by the method of least squares to the simpler equation

$$G_{xp}^E = ax_1x_2;$$

the value of a was found to be +414 cal./mol.

The consistency of the data was tested by plotting experimental values of $(\mu_1^E - \mu_2^E)$ against x_1 and measuring the areas above and below datum. This ratio was found to be 1.00.

(c) *The System n-Hexane-Chlorobenzene at 760 mm. Hg*

The equilibrium data for this system were measured at 759.8 ± 0.4 mm. Hg to enable a comparison to be made with results obtained using other equilibrium

TABLE 8
n-HEXANE-CHLOROBENZENE AT 65 °C.

x_1	y_1	P (mm. Hg std.)	μ_1^E (cal./mol.)	μ_2^E (cal./mol.)	G_{xp}^E (cal./mol.)	α	ΔG_{xp}^E (cal. mol.)
0.083	0.544	166.4	344	2.1	30.4	13.2	-1.0
0.144	0.679	222.3	315	2.5	47.4	12.6	-3.6
0.201	0.744	264.1	265	11.6	62.5	11.5	-4.0
0.284	0.803	319.7	210	35.8	85.3	10.2	+1.1
0.394	0.852	382.3	148	69.8	100	8.86	+1.6
0.438	0.866	403.9	123	92.1	106	8.27	+3.9
0.485	0.882	428.3	106	102	104	7.93	+0.6
0.540	0.896	453.8	81.3	133	105	7.30	+2.4
0.591	0.910	477.9	65.4	143	97.3	7.02	-2.8
0.679	0.929	516.3	36.6	198	88.6	6.19	-1.7
0.806	0.957	578.0	14.1	270	63.7	5.37	-1.0
0.927	0.984	638.4	2.7	345	27.6	4.72	-0.3

stills. The pressure was controlled by a modification of the manostat described by Gould and Evans (1933). The results are given in Table 9. These data were shown to be consistent by the method of Herington (1951).

 TABLE 9
n-HEXANE-CHLOROBENZENE AT 759.8 ± 0.4 MM. Hg

x_1	y_1	P (mm. Hg)	t (°C.)	μ_1^E (cal. mol.)	μ_2^E (cal. mol.)	α
0.018	0.118	759.7	127.56	376	1.3	7.45
0.049	0.282	759.4	121.06	353	4.0	7.62
0.081	0.406	759.6	115.66	328	2.7	7.74
0.109	0.491	759.9	111.53	310	0.3	7.87
0.146	0.577	759.8	106.62	289	3.9	7.96
0.200	0.666	759.8	101.04	255	6.0	7.97
0.309	0.769	759.7	92.70	184	42.1	7.42
0.419	0.835	759.7	86.84	129	70.5	6.98
0.516	0.872	759.7	82.66	90.3	120	6.38
0.591	0.896	759.7	80.19	65.0	162	5.94
0.593	0.896	760.0	80.13	62.6	160	5.89
0.594	0.896	760.1	80.17	61.0	156	5.91
0.644	0.912	759.8	78.31	47.7	183	5.70
0.737	0.934	760.0	75.70	23.6	256	5.06
0.790	0.950	760.1	74.17	16.3	255	5.09
0.793	0.950	760.4	74.14	14.3	278	4.96
0.847	0.965	759.4	72.72	5.9	276	4.91

(d) *The System cycloHexane-Nitrobenzene at 80° C.*

This system was chosen to test the performance of the still with a mixture of very high relative volatility. The results are given in Table 10. The excess

free energy data were fitted by the method of least squares to the equation that was used for the system *n*-hexane-chlorobenzene; the value of *a* was found to be 1630 cal./mol. The difference between the experimental and fitted values of G_{xp}^E shown in the last column of Table 10 are less than the estimated errors in G_{xp}^E of approximately 60 cal./mol.

TABLE 10
cyclohexane-nitrobenzene AT 80 °C.

x_1	y_1	P (mm. Hg)	μ_1^E (cal./mol.)	μ_2^E (cal./mol.)	G_{xp}^E (cal./mol.)	α	ΔG_{xp}^E (cal./mol.)
0.145	0.973	322.6	490	130	182	213	-20
0.440	0.985	556.2	369	380	375	83.6	-27
0.490	0.985	573.1	314	465	391	68.4	-16
0.573	0.985	600.3	236	606	394	50.0	-5
0.686	0.985	630.3	143	867	371	30.1	+19
0.773	0.983	654.1	83	1210	339	17.0	+52
0.891	0.991	689.9	25	1290	162	13.9	+4

The upper limit of relative volatility for satisfactory performance of the still could not be defined by these results as the relative volatility was too high over most of the composition range. The upper limit for α is somewhere in the range covered by this system and is probably close to 14.

Difficulty was experienced in operating the still at low *cyclohexane* concentration due to the large differences in vapour pressure and density of the components. The condensed vapour returned to the kettle unevenly giving irregular boiling.

VI. CONCLUSIONS

The results that have been given for these three binary systems show that the new equilibrium still gives consistent data over a wider range of relative volatility than still No. 1. Thermodynamically consistent results can be obtained with the apparatus described for miscible systems having a relative volatility from 1.0 to 14. The still is satisfactory for obtaining accurate data at ordinary pressures both for thermodynamic investigation of liquid mixtures and for plant design.

VII. ACKNOWLEDGMENTS

The author thanks all those who helped with this work, particularly Miss P. C. Miller for curve fitting and Messrs. O. H. Rigby, G. English, L. N. Sutton, and F. W. Wheatley for their help in the purification of the components and in the equilibrium measurements; also Mr. D. T. C. Gillespie for suggestions during the development of the apparatus.

VIII. TABLE OF NOMENCLATURE

A, B, C, = constants in the Antoine equation (1),

a, b, c, = constants in equations (6), (7), and (8),

G =free energy (Gibbs),

n_D^{25} =refractive index at 25 °C. for sodium D line,

P =pressure,

P_1, P_2 =vapour pressures of components 1 and 2,

R =gas constant=1.9872 cal./deg. mol.,

T, t =temperature °K., °C. (0 °C.=273.16 °K.),

V =molar volume,

d_4^{25} =density at 25 °C. referred to water at 4 °C.,

f =constant for converting from mm. Hg–litres to calories=0.0318657,

x, y =mole fraction in liquid and vapour respectively,

α =relative volatility,

β =second virial coefficient in equation of state,

γ =activity coefficient,

$\delta()$ =absolute error in (),

$\Delta()$ =difference between fitted and experimental values of (),

μ =chemical potential,

Superscript

E =excess change above that for an ideal solution,

Subscripts

1, 2 refer to the more and less volatile components respectively,

x indicates a function of composition,

p indicates mixing at constant pressure,

c refers to the critical state.

IX. REFERENCES

- BROWN, I., and EWALD, A. H. (1950).—*Aust. J. Sci. Res. A* **3**: 306.
 BROWN, I., and EWALD, A. H. (1951).—*Aust. J. Sci. Res. A* **4**: 198.
 EGLOFF, G. (1939).—"Physical Constants of Hydrocarbons." Vol. 1. p. 40. (Reinhold: New York.)
 GOULD, F. A., and EVANS, J. C. (1933).—*J. Sci. Instrum.* **10**: 215.
 HERINGTON, E. F. G. (1951).—*J. Inst. Petrol.* **37**: 457.
 INTERNATIONAL CRITICAL TABLES (1928).—Vol. 3. (McGraw-Hill: New York.)
 MEISSNER, H. P. (1949).—*Chem. Engng. Prog.* **45**: 149.
 REDLICH, O., and KISTER, A. T. (1948).—*Industr. Engng. Chem.* **40**: 345.
 ROSSINI, F. D., *et al.* (1947).—Selected values of properties of hydrocarbons. Circ. U.S. Nat. Bur. Stand. C.461.
 SHEPARD, A. F., and HENNE, A. L. (1930).—*Industr. Engng. Chem.* **22**: 356.
 TIMMERMANS, J. (1950).—"Physico-chemical Constants of Pure Organic Compounds." (Elsevier: New York.)
 TOMPA, H. (1948).—*J. Chem. Phys.* **16**: 292.
 WILLINGHAM, C. B., TAYLOR, W. J., PIGNOCCO, J. M., and ROSSINI, F. D. (1945).—*J. Res. Nat. Bur. Stand.* **35**: 219.

AN INVESTIGATION OF ERRORS IN THE AMPEROMETRIC AND STARCH INDICATOR METHODS FOR THE TITRATION OF MILLI-NORMAL SOLUTIONS OF IODINE AND THIOSULPHATE

By J. H. BRADBURY* and A. N. HAMBLEY†

[Manuscript received April 30, 1952]

Summary

The amperometric method of Foulk and Bawden when applied to the titration of 0.001N thiosulphate solutions with iodine is shown to be more precise and to have a smaller "blank" than that using starch as indicator. With simple precautions the method has an accuracy of at least 0.1 per cent. over the range pH 1-8, but at higher pH there is serious error due to sulphate formation. The magnitudes of other errors are determined and it is shown that all can be rendered negligible in titrations of milli-normal solutions.

I. INTRODUCTION

The method of titration in which the end-point is detected by the sudden increase, or the sudden decrease, of current flow between two similar electrodes immersed in the solution, due to polarization effects, has received various names. The names "dead-stop end-point method" (Foulk and Bawden 1926) and "depolarimetry" (Guzman and Rancano 1934; Guzman 1935) are applicable to the titration performed in a particular direction; the first to the case in which the end-point is marked by the polarization of one, or both, electrodes and the second to that in which it is marked by the depolarization of the electrodes. Salomon (1897*a*, 1897*b*), who first performed titrations of this type, called them "galvanometric titrations". They are, however, merely a special case of the amperometric titration in which the change in the value of the current at the end-point is so extreme that there is no need to plot values on a graph. We shall show in a subsequent paper that the end-point can also be found, as in other amperometric titrations, by the intersection of two linear graphs of galvanometer deflection, just before and just after this end-point, *versus* volume of reagent added. We shall therefore refer to the method as an amperometric titration (cf. Kolthoff and Lingane 1941).

A preliminary study of the amperometric titration of thiosulphate solutions with iodine, first described by Foulk and Bawden (1926), showed that this method, particularly with dilute solutions (10^{-3} to 10^{-4} N) had a more easily detected end-point, free from subjective errors, and required a smaller excess of free iodine than the classical starch indicator method. It was therefore considered that it would provide a very sensitive method for detecting departures

* Present address: Chemistry Department, University of Birmingham.

† Chemistry Department, University of Melbourne.

of the volumetric measurements from the stoichiometric relationship of the titration reaction, which is assumed to be



The possible causes of a departure of the indicator end point from that required by the above equation are: (a) instability of (i) thiosulphate and (ii) iodine solutions; (b) hydrolysis of iodine to hypiodous hypiodite or oxidizing thiosulphate at the pH of the titration; (c) the local consumption of excess iodine required to depolarize the cathode; (d) adsorption of iodine on glass surfaces; (e) oxidation of (i) thiosulphate or (ii) tetrathionate to sulphuric acid; (f) reaction of intermediates such as $S_4O_6^{2-}$ to reach completely before the indicator end point; (g) aerial oxidation of iodide to iodine at low pH; and (h) aerial oxidation of iodine to iodate at high pH (Kolthoff and Sandell 1956).

Effects due to all of these except (d) and possibly (f) have been shown in our experiments. Aerial oxidation of iodine to iodate has only been observed in iodometric oxidation of arsenite, in arsenite at pH values outside the range for which accurate values can be obtained in the corresponding oxidation of thiosulphate to tetrathionate.

11. REAGENTS

Water.—Distilled water was redistilled in an all glass system, with the rejection of first and last fractions, to give a product with conductivity $0.85 \pm 1.1 \times 10^{-7}$ ohm⁻¹ cm.⁻¹. This water was boiled, cooled, and stored through a G4 sintered glass plate immediately before making the standard solutions.

Iodine.—Reagent "resublimed iodine" was sublimed at 90°C. from admixture with potassium iodide and the product again sublimed at the same temperature without addition of potassium iodide. The purified reagent was stored over anhydrous calcium chloride.

Arsenious Oxide.—"AnalAR" arsenious oxide was twice sublimed under reduced pressure and the product was shown to be free from arsenate by standard tests (Rosin 1946).

Sodium Thiosulphate.—Merck, A.R. was twice crystallized from water at 65°C. and the pentahydrate dehydrated at 120°C. to constant weight (Le Roy and Thomsen 1931; Vendelsson and Skogstad 1941). For preparation of solutions in which thiosulphate was not a primary standard the recrystallized pentahydrate was used.

Potassium Iodide.—Potassium iodide (A.R.) was tested and found to contain less than 26 parts per million of iodine and to give less than 1 part per million of free iodine in a freshly prepared 0.1N solution.

Buffer Solutions.—These were prepared as 0.2M solutions from concentrated potassium hydrogen phthalate, sodium dihydrogen phosphate, and borax and were adjusted to the desired pH with 0.1N hydrochloric acid, or sodium hydroxide and checked with a Radiometer glass electrode pH meter. Other

treated with potassium iodide and sulphuric acid and titrated at pH 2.5 with iodine to the starch end-point the blank determinations were in good agreement and showed the absence of substances capable of reducing iodine at this pH.

Starch.—"AnalaR" starch was mixed to a paste with cold water and treated with boiling water to give a 0.25 per cent. solution. Fresh solution was prepared every second day.

III. APPARATUS

All volumetric apparatus was calibrated in the usual manner. Amperometric titrations were performed in a 300 ml. three necked flask. The central neck carried the stirrer and the side necks the pair of electrodes and the outlet of the burette respectively. The 10⁻³N iodine was stored in a litre reservoir connected to a bulb burette from which 0.5 and 17.5–22.5 ml. could be delivered with an accuracy of ± 0.002 ml. As solution was removed from the reservoir, it was replaced by air that had passed through two 100 ml. bubblers in series, each containing 10⁻³N iodine. The iodine was delivered below the surface of the liquid being titrated.

A potential of 10–20 mV. was applied to the electrodes by means of an accumulator and voltage divider. The current flowing between them was measured in arbitrary units by the deflections of a sensitive galvanometer (c. 1500 mm. μ A.).

Starch end-points were observed in a Hellige comparator fitted with two 4 cm. cells, one containing distilled water and the other the solution to be titrated. Constant conditions of illumination were maintained.

Sulphate determinations were made in a Pulfrich nephelometer (Zeiss) which was calibrated by means of solutions 5.0×10^{-7} to 2.0×10^{-5} molar in sulphate. To 47 ml. of the sulphate solution was added 0.5 ml. of 1M hydrochloric acid and 2.5 ml. of 1M barium chloride solution and the scattered light observed after an interval of 1 hour. The reproducibility of the calibration varied from 50 per cent. of the measured concentration at the lower end of the range to 3 per cent. at the upper end.

IV. RESULTS AND DISCUSSION

(a) *Stability of Solutions*

(i) *Stability of Sodium Thiosulphate and its Solutions*.—It has been claimed (La Mer and Tomlinson 1937; Tomlinson and Ciapetta 1941) that recrystallized samples of $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$ undergo decomposition on standing for a few months at room temperature with formation of sulphur and sodium sulphite. Determinations of sulphite (plus sulphate) in various samples were made by dissolving about 0.5 g. in water, adding first 0.1N iodine, and then 10⁻³N iodine to the amperometric end-point. Standard amounts of hydrochloric acid and barium chloride were then added to precipitate barium sulphate. The values obtained (Table 1) have an accuracy of about 10 per cent. and include any sulphate formed by oxidation of thiosulphate by iodine.

Anhydrous sodium thiosulphate prepared by the method of Tomlinson and Ciapetta (1941) contains slightly larger amounts of sulphite plus sulphate than the hydrate from which it is prepared, but the impurity is still much less

TABLE 1
SULPHATE CONTENT OF SODIUM THIOSULPHATE AND ITS SOLUTIONS

Description of Thiosulphate	Sample Number	Time of Standing (months)	Mole of ($\text{SO}_3^- + \text{SO}_4^-$) per 100 Mole $\text{S}_2\text{O}_3^{2-}$
Pentahydrate (Merck, A.R. once recrystallized from boiling H_2O)	1	8	0.023
Pentahydrate (sample 1 three times recrystallized from H_2O at 65°C .) ..	2	2.5	0.015
Pentahydrate (sample 1 once recrystallized from H_2O at 65°C .)	3	0.5	0.015
Anhydrous (sample 2 dried at 120°C . for 3 days)	4	2.5	0.035
Anhydrous (sample 3 dried at 120°C . for 2 days)	5	0.5	0.030
0.1N solution (prepared from sample 1 and kept 3.5 months)	6	3.5	0.045

than one part per thousand and the anhydrous substance is a satisfactory standard (Young 1904) for work with an accuracy of 0.1 per cent. (Table 2). Titrations were performed, on aliquots of thiosulphate diluted to c. 60 mL. by

TABLE 2

Primary Standard	Gram Equivalents I_2 per Kilogram Solution		
	Solution 1	Solution 2	Solution 3
I_2	0.09633	0.09631	0.09636
$\text{Na}_2\text{S}_2\text{O}_3$ (anhydrous)	0.09633	0.09633	0.09636
As_2O_3	0.09626		

addition of the standard iodine solution (0.1N) from a weight pipette until the end-point was nearly reached and then the titration completed to the amperometric end-point with 10^{-3}N iodine. The titration of arsenite was performed similarly but in a solution buffered at pH 8.2.

Decinormal sodium thiosulphate solutions prepared in freshly boiled and cooled, redistilled water undergo only slight decomposition over a period of

100 days (Table 1) and the titre determined periodically with freshly prepared 0.1N iodine did not change by 0.1 per cent. within 1 month. Millinormal solutions prepared by further dilution of the decinormal solution were shown by a similar method to be stable within the same limits for a period of at least 30 hours.

(ii) *Stability of Iodine Solutions.*—Decinormal iodine solutions prepared by dissolving 0.1 equivalent of resublimed iodine in a freshly prepared, concentrated, aqueous solution of 0.25 equivalents of potassium iodide and making up to 1 litre are found to be quite stable if kept in a blackened bottle, provided precautions are taken to avoid volatilization losses if the bottle is frequently opened. Solutions containing 0.1 equivalent of iodine and 0.1 equivalent of potassium iodide per litre are satisfactory for immediate use but show somewhat larger losses by volatilization. Table 2 shows that no analytically appreciable loss occurs in making up these solutions.

Millinormal iodine solutions prepared by dilution of 0.1N iodine, 0.1N potassium iodide from a weight pipette delivering below the surface of a measured volume of water in the reservoir of the amperometric apparatus showed a consistent "loss of iodine" of about 0.5 per cent. (10^{-7} equivalents I_2 in a 20 ml. titration) when compared with freshly prepared 0.001N thiosulphate solution. If a further quantity of 0.1N iodine is added to a solution in which the "loss" has occurred, the increase in titre due to the second addition of iodine is quantitative. The amount of possible loss by volatilization into the free space above the solution can be calculated to be negligibly small. The loss is not increased by increase of the area of glass in contact with the solution from 500 to 1500 cm.² (Table 3) and is therefore not due to adsorption. No conversion of hypiodite (formed by hydrolysis of iodine) to iodate has occurred since the loss was the same when the titrations were performed at pH 8, 5, and 1.5. Addition of a known quantity of iodate made no difference to the titre at pH 8 or 5 but gave a quantitative increase at pH 1.5. Treatment of the distilled water by boiling with alkaline permanganate followed by two further distillations made no change. On the other hand, if 0.001N solution was prepared by diluting approximately 5 g. of 0.1N solution to 500 ml. with the distilled water and then the whole was titrated first with 0.1N and finally with 0.001N thiosulphate to the amperometric end-point, the mean loss on 15 such dilutions was found to be only 0.06 per cent. In these titrations where thiosulphate is added to iodine, the iodide concentration was 0.001N or greater throughout the titration. It was then found that an increase in the iodide concentration in the 0.001N iodine solution reduced the amount of "loss" for the reverse titration, and since the stability of the iodine solutions was found to be increased by the presence of acid, a series of dilutions was performed with various additions of acid and iodide (Table 3). It is apparent that solutions prepared to be 0.01N with respect to acid and 0.02N with respect to iodide show negligible "dilution loss". Such solutions maintain their titre for a period of 15 hours. Solutions prepared at lower acidity and iodide concentration show small but appreciable

variations in titre during the first 5 hours but then maintain their titre for about 15 hours. Millinormal iodine solutions should always be standardized after dilution and not credited with the normality calculated from the dilution ratio.

TABLE 3
EFFECT OF IODIDE AND ACID ON IODINE "LOSS"

Acid Concentration		Iodide Concentration (I ⁻)	Mean Loss of Iodine (%)
HCl	H ₂ SO ₄		
Nil	Nil	0.0010	0.48
Nil	0.0010	0.0010	0.45
Nil	Nil	0.0025	0.28
Nil	0.0010	0.0025	0.21
Nil*	0.0010	0.0025	0.21
0.0010	Nil	0.0025	0.23
0.0010†	Nil	0.0025	0.22
0.010	Nil	0.0025	0.15
Nil	Nil	0.020	0.10
0.001	Nil	0.020	0.11
0.010	Nil	0.020	0.03

* Water redistilled from alkaline permanganate used.

† Area of glass in contact with solution increased from 500 to 1500 cm.²

by addition of small pieces of cleaned Pyrex tubing.

(b) Hydrolysis of Iodine

The equilibrium constant for the hydrolysis of iodine,



has been shown to be $3 \cdot 10^{-13}$ at 25 °C. (Bray 1910, 1911). Since a small finite concentration of free iodine is required to indicate the end-point of the titration either with the amperometric or starch indicator method, it is expected that the excess of solution required to perceive the end-point should increase with pH at a given iodide concentration. The hydrolysis constant, however, indicates that the fraction hydrolysed will be extremely small and should only increase by about 4 per cent. between pH 5 and 9. *Direct* "blank" determinations were carried out by addition of 10^{-3}N iodine solution to 100 ml. of 0.01M buffer solution containing a known iodide concentration, with an applied potential of 20 mV. between the electrodes. The titration was carried to the first permanent increase of current between the electrodes. *Indirect* determinations were made by using a solution as above together with 1 ml. of 0.001N thiosulphate solution. After the first end-point was reached a second ml. of thiosulphate solution was added and a second end-point determined (Evans and Simmons

1944; Evans 1947). The difference between the two volumes of iodine used gives the "end-point blank" which will include any effects due to the presence of thiosulphate and tetrathionate during the titration. The values of the "blank" obtained have no absolute significance since they vary from one pair of electrodes to another and even for the same pair of electrodes they change slightly during a series of determinations. To eliminate such minor variations the same pair of electrodes was used throughout and the experimental series were designed so that a standard determination was repeated at intervals during each set and small corrections made for the very slight variation in the standard value. The tabulated results therefore show only effects due to variation in solution composition. For comparison, direct blanks were determined using starch indicator, the titration being performed in a 4 cm. cell in a comparator.

TABLE 4

AMPEROMETRIC AND STARCH-INDICATOR BLANKS IN ML. 0.001N IODINE PER 100 ML. OF SOLUTION
AT END-POINT FOR VARIOUS VALUES OF pH AND IODIDE CONCENTRATION

pH	Blank at $[I^-]=8 \times 10^{-3}$			$[I^-]$	Blank at pH 5	
	Starch	Amperometric			Amperometric	
		Direct	Indirect		Direct	Indirect
3.0	0.210	0.026	0.022	9.7×10^{-2}		0.280
5.1	0.246	0.032	0.032	4.85×10^{-2}	0.140	0.168
6.1	0.218	0.034	0.034	1.94×10^{-2}		0.060
7.0	0.220	0.036	0.036	9.7×10^{-3}	0.040	0.046
7.5	0.248	0.038	0.028	4.85×10^{-3}	0.036	0.040
8.0	0.260	0.040	0.046	8×10^{-4}	0.020	0.032
8.6	0.362	0.054	0.026	4.85×10^{-4}	0.022	0.028
9.1	0.390	0.092	0.038	5×10^{-5}	0.016	0.020

The values recorded in Table 4 are each mean values from three measurements. It is apparent that, while the blank by the indirect amperometric method varies only slightly with pH, both the blank by the direct method and that using starch indicator increase rapidly at pH 8.6 and 9. The increase in the case of starch is presumably due to slight oxidation of the indicator in the alkaline solution, but we cannot account for the similar increase shown in column 3, which, while in the expected direction, is too great in magnitude. It is possible that the figures in column 4, at pH greater than 8, are not related to those in column 3 for, as is shown in Section III (e), titrations of thiosulphate under these conditions have an error due to sulphate formation. The starch blanks are six to eight times the size of those obtained by the amperometric method and also show a much greater variance in triplicate determinations despite the attempt to standardize the conditions for their observation.

The small influence of iodide ion in suppressing hydrolysis of iodine is completely obscured by its reaction with iodine to form the tri-iodide ion. At pH 5 even with the lowest iodide concentration recorded in Table 4 the hydrolysis of iodine is calculated to be negligibly small.

(c) *Free-Iodine Concentration to give Indicated End-Point*

From the known value of the equilibrium constant for tri-iodide formation at 18 °C. (which was the temperature of our experiments)

$$K = \frac{[\text{I}_2][\text{I}^-]}{[\text{I}_3^-]} = 1.17 \times 10^{-3},$$

we can calculate the free-iodine concentration at the end-points for various iodide concentrations.

If $[\text{I}^-]_t = [\text{I}^-] + [\text{I}_3^-] \simeq [\text{I}^-]$ when $[\text{I}_3^-] \ll [\text{I}^-]$ as in all cases shown, and $[\text{I}_2]_t = [\text{I}_2] + [\text{I}_3^-]$ (all concentrations being expressed in mole/l.) then we obtain

$$[\text{I}_2] = \frac{1.17 \times 10^{-3} [\text{I}_2]_t}{[\text{I}^-]_t + 1.17 \times 10^{-3}}$$

Using this expression, the values of $[\text{I}_2]$ shown in Table 5 are obtained from the indirect amperometric "blanks" as the free-iodine concentration required to indicate the end-point.

TABLE 5
CALCULATED FREE-IODINE CONCENTRATION AT AMPEROMETRIC END-POINT
FROM BLANKS BY INDIRECT METHOD

$[\text{I}^-]$ (mole/l.)	Blank in ml. 10 ⁻³ N Iodine per 100 ml. Solution	$[\text{I}_2]_t$ (mole/l.)	$[\text{I}_2]$ (mole/l.)
1.94×10^{-1}	0.556	2.78×10^{-6}	1.93×10^{-6}
9.7×10^{-2}	0.280	1.40×10^{-6}	1.67×10^{-6}
4.85×10^{-2}	0.168	0.84×10^{-6}	1.98×10^{-6}
1.94×10^{-2}	0.060	0.30×10^{-6}	1.71×10^{-6}
9.7×10^{-3}	0.046	0.23×10^{-6}	2.47×10^{-6}
9.9×10^{-4}	0.040	0.20×10^{-6}	1.09×10^{-7}
1.20×10^{-4}	0.030	0.15×10^{-6}	1.36×10^{-7}

The values found for iodide concentrations greater than 1×10^{-2} N indicate that a constant free-iodine concentration is required to depolarize the cathode. The apparent increase at low iodide concentration is discussed in Section III (d) below. For comparison the blanks with starch indicator are treated similarly in Table 6. These show the increase of free iodine required to give the blue colouration as the iodide concentration decreases, in accord with the well-known dependence of this reaction on iodide concentration.

The results for the amperometric blanks are in marked contrast with those of Evans (1947), who claims that they decrease to zero at an iodide concentration of $2.4 \times 10^{-2}N$, whereas we find that they increase steadily with iodide concentration, due to tri-iodide formation. We can only attribute his result to the

TABLE 6
CALCULATED FREE-IODINE CONCENTRATION AT STARCH END-POINT

[I ⁻] (mole l.)	Blank in ml. 10 ⁻³ N Iodine per 100 ml. Solution	[I ₂] _t (mole/l.)	[I ₂]* (mole/l.)
9.7×10^{-2}	0.328	1.64×10^{-6}	2.0×10^{-8}
9.7×10^{-3}	0.146	0.73×10^{-6}	7.9×10^{-8}
9.7×10^{-4}	0.154	0.77×10^{-6}	4.2×10^{-7}
1.0×10^{-4}	0.302	1.51×10^{-6}	1.4×10^{-6}

* No allowance has been made for adsorption on starch or glass.

presence of a trace of iodine in his 10 per cent. iodide solution (see Section III (h)). Evans argues that a slight excess of iodine can overcome the polarization of the anode, due to the presence of insufficient iodide. This will only be the case if the hydrolysis of the excess iodine provides the necessary iodide concentration. We have found, by an indirect method, that the minimum concentration of iodide required to depolarize the anode is $c. 1.5 \times 10^{-6}$ mole/l. As this concentration is exceeded in all the determinations, we do not experience any trouble due to this cause and would not expect it in any of Evans's titrations in which the minimum iodide concentration at the end-point was $1.7 \times 10^{-5}N$.

(d) Adsorption of Iodine on Glass

Gilbert and Marriott (1948) record a loss of $c. 4 \times 10^{-8}$ mole of iodine on addition of a solution to 800 ml. of a 0.001N iodide solution at pH 6 in a 1 l. round-bottomed flask. They attribute this loss to adsorption on the glass surface. The conditions of their experiment are very similar to those used in the determination of amperometric blanks by the direct method and it is easily shown that the effects of such an adsorption would be to increase the calculated free-iodine concentration required at the end-point when the iodide concentration is low (Table 5).

If adsorption occurs, then we have

$$[I_2] - (I_2)_{ad} + [I_3^-] = [I_2]_t$$

$$[1 - [I^-]_t K] [I_2] - (I_2)_{ad} = [I_2]_t$$

where $K = 1.17 \times 10^{-3}$. Using the various values of $[I^-]$, and $[I_2]_t$ for the six determinations of the average blank (direct method) in Table 7 we obtain six equations for the two unknowns $[I_2]$ and $(I_2)_{ad}$. From these, by a "least-squares" method one obtains 1.46×10^{-8} mole/l. as the most probable value of

the iodine concentration required to depolarize the cathode and 8.15×10^{-9} mole as the amount of iodine adsorbed from the 100 ml. of solution onto the glass surface. From the values of Δb (Table 7), the error in the blank calculated from these figures, it is seen that they account adequately for the measured blanks.

TABLE 7

CHECK ON AMPEROMETRIC BLANKS AFTER ALLOWING FOR A CRITICAL CONCENTRATION OF FREE IODINE AND FOR ADSORPTION IN GLASS

$[I^-]_t$	Blank in ml. $10^{-3}N$ Iodine per 100 ml. of Solution	$[I_2]_t$	$[I_2]_{calc}$ (no adsorption)	Δb (ml.)
4.85×10^{-2}	0.140 ± 0.000	7.0×10^{-7}	1.65×10^{-8}	0.00
9.70×10^{-3}	0.040 ± 0.002	2.0×10^{-7}	2.15×10^{-8}	-0.003_4
4.85×10^{-3}	0.036 ± 0.004	1.8×10^{-7}	3.50×10^{-8}	$\pm 0.004_6$
4.85×10^{-4}	0.022 ± 0.002	1.1×10^{-7}	7.8×10^{-8}	$\pm 0.001_8$
9.7×10^{-5}	0.019 ± 0.002	0.95×10^{-7}	8.8×10^{-8}	-0.000_4
4.9×10^{-5}	$0.016_4 \pm 0.002$	0.82×10^{-7}	7.9×10^{-8}	-0.002_8

From the dimensions of the apparatus the adsorption is calculated to be 7.8×10^{-11} mole/cm.², while if we allow for a concentration of 1.5×10^{-8} mole/l. to make the electrodes function reversibly in Gilbert and Marriott's experiment, the adsorption which they noted can be calculated to be 2.5×10^{-11} mole/cm.².

Titrations carried out in the presence of, and in the absence of, Pyrex glass powder show small but significant differences confirming the reality of the adsorption phenomenon.

(c) Errors due to Sulphate Formation

(i) *Oxidation of Thiosulphate to Sulphate.*—The most serious source of error in the titration of thiosulphate with iodine arises through the oxidation of thiosulphate to sulphate



The reaction will obviously be more important at high pH and high oxidation potential of iodine, i.e. low iodide concentration and high iodine concentration. Thus it is possible to oxidize thiosulphate quantitatively to sulphate by means of iodine in the presence of normal sodium hydroxide (Abel 1912; Kolthoff 1921; Frost 1944).

To test the dependence of the titre of iodine on the pH of the solution it was assumed with Kolthoff (1921) that titrations of millinormal solutions at pH 5 are free from error due to sulphate formation. Alternate titrations were carried out at the desired pH and at pH 5, so that allowance could be made for any slow alteration of solutions with time and for any modification of the sensitivity of the electrodes. To the 0.001N thiosulphate solution (20 ml.)

was added 5 ml. of buffer solution diluted with an equal volume of water and the titration carried out to the amperometric end-point. An aliquot (25 ml.) of this solution was then transferred to the comparator cell, 2 ml. of 0.25 per cent. starch solution added, and the titration carried on with 0.001N iodine to the first perceptible blue colour. The total iodine required between the two end-points was then calculated from the ratio of the volume at the amperometric end-point to 25 ml. Since the iodide concentration rose from zero to its final value, only the latter is recorded in Table 8. The pH also varied slightly during titration due to the presence of stabilizing acid in the iodine solution: the initial and final values are recorded. At each pH the appropriate amperometric or starch "blank" (Table 5) was deducted from the titre to give the equivalent time and the values of the mean differences (from triplicate determinations, between the titrations at various pH values and that at pH 5 are shown in the table as percentage errors.

TABLE 8

PERCENTAGE ERRORS IN TITRATIONS OF 0.001N THIOSULPHATE AND ARSENITE WITH IODINE AT VARYING pH

pH	Error (%) for Thiosulphate			Error (%) with Arsenite [I ⁻]=9.3/10 ⁻³ M
	[I ⁻]=8.0/10 ⁻⁴ M	[I ⁻]=9.3/10 ⁻³ M	[I ⁻]=9.3/10 ⁻³ M	
	Starch	Amperometric	Amperometric	
0.8-1.1			-0.012	
1.3-1.5			0.006	
2.0			0.003	
3.0	0.017	-0.004		
5.1-4.9	0.000	0 (definition)	0 (definition)	
6.1-5.8	0.020	0.016		
7.0-6.9	0.028	0.003		0 (definition)
7.5-7.4	0.029	-0.002		
8.0-7.8	0.031	0.014		
8.6-8.5	0.121	0.108	0.019	-0.004
9.1-9.0	0.367	0.345	0.196*	-0.006

The starch reduction values are not as reliable as those from amperometric measurement, yet they confirm that significant sulphate formation does not occur below pH 8. The amperometric values also show that the error due to sulphate formation at pH greater than 8 is reduced by increased iodide concentration. The observation that the titration of arsenite with iodine shows no significant change between pH 7 and 9 shows that the effect is not due to an impurity in the borax buffer. A microchemical determination of sulphate in the experiment indicated by an asterisk in Table 8 gave 3.7×10^{-7} mole compared

to 4.9×10^{-8} mole deduced from the increased titre. Kolthoff (1921) has indicated that accurate titrations of 0.1N iodine into thiosulphate can be performed in the presence of borax buffers; this does not apply with millinormal solutions although the titration is satisfactory at a pH less than 8.

(ii) *Oxidation of Tetrathionate*.—Dodd and Griffith (1949) have made some rate studies of the oxidation of tetrathionate to sulphate by iodine, but their rate "constants" are not suitable for the calculation of the reaction of tetrathionate with small excesses of iodine at the iodide concentrations used in our work. The reaction is, however, extremely slow compared with the reaction of thiosulphate with iodine, so that errors would only be expected if excess iodine is allowed to stand in contact with tetrathionate, as in cases where a titration of an iodine solution is carried nearly to the end-point with say 0.1N thiosulphate and then completed with millinormal reagent. During such titrations, losses of the order of 0.3 ml. of 0.001N iodine (mean of 4 determinations) have been found when 500 ml. solution $[I_2] = c. 10^{-5}M$ was allowed to stand in a closed vessel for 2 hours between the two parts of the titration, but under conditions in which the volatilization loss could be calculated to be quite negligible. No appreciable error was found if the titration was completed within 15 minutes.

(f) *Formation of Intermediates such as $S_2O_3I^-$*

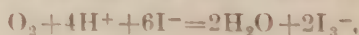
Raschig (1915) postulated that the reaction of thiosulphate with iodine proceeded by the two stages:



in order to account for the thiosulphate catalysed reaction between iodine and sodium azide. Kolthoff (1921) found that the blue starch iodine colour is re-formed within a few seconds of its removal by the rapid addition of thiosulphate-starch solution to 0.001N iodine, a behaviour that can be readily explained if some $S_2O_3I^-$ was present at the first end point. Although Dodd and Griffith (1949) have claimed to measure the amounts of $S_2O_3I^-$ present in solutions under various conditions, Abel (1950) has indicated that there is an alternative explanation of their results without recourse to this hypothetical intermediate. The only possible evidence for such an intermediate in our study comes from a comparison of the blanks obtained by the direct and indirect methods (Table 4, columns 6 and 7) and a number of similar tests at various iodide concentrations. A statistical test indicates that the blank obtained by method 2 is significantly greater than that obtained by method 1 (except at high pH). These small and analytically negligible differences may however be due to a slightly greater oxidation to sulphate in the first ml. of the titration than in the second ml. In the latter the oxidizing potential of local concentrations of iodine will be reduced by the higher iodide concentration.

(g) *Oxidation of Iodide by Air*

The aerial oxidation of iodide will occur mainly by the reaction



and will be favoured by high acidity and high iodide concentration. We have found that a molar solution of potassium iodide prepared in redistilled water undergoes slow oxidation (cf. Polissar 1935). After 3 days' storage in a stoppered bottle the iodine concentration determined by titration of an aliquot with 0.001N thiosulphate at pH 5 was 4×10^{-6} N and after 10 days 3.5×10^{-5} N. Concentrated solutions of potassium iodide such as are used in the preparation of standard solutions from resublimed iodine (Kolthoff and Sandell 1936, p. 592) should therefore be made immediately before use.

To test the possible influence of liberation of iodine from iodide by aerial oxidation during a titration, a solution was prepared containing 1×10^{-5} N thiosulphate, 0.02N iodide, and acidified to pH 1, i.e. using much higher iodide and hydrogen ion concentrations than are present in recommended procedures. By titration with 5×10^{-4} N iodine solution it was found that the thiosulphate decreased in concentration by 3×10^{-7} equivalents per litre (mean of 3 determinations) when the solution was stirred for 10 minutes in the amperometric apparatus. This would correspond to an error of only 0.3 part per thousand in the titration of 0.001N thiosulphate even with these extreme conditions, so that no error from this source should arise in our general titration procedure.

(b) *Comparison of the Precision of the Amperometric and Starch-Indicator Methods*

Equivalent titres of thiosulphate and iodine were determined by titration of 20 ml. aliquots of 0.001N thiosulphate with 0.001N iodine in the amperometric apparatus to the current end-point. A 25 ml. aliquot of this solution was then transferred to the comparator cell and the titration carried to the starch end-point as described in Section III (c). The "starch titre" was then calculated from the ratio of the volume of solution at the amperometric end-point to 25 ml. The end-point blanks for the known iodide concentrations and pH were then subtracted. As is obvious from this procedure, the amperometric end-point requires a smaller excess of iodine than does the starch end-point. It remains to compare the precision with which the two end-points can be reproduced. With 154 such titrations (usually carried out in groups of three) a statistical test showed that there was less than an 0.1 per cent. chance that the starch titrations were as precise as the amperometric titrations. This lesser precision occurs despite the standardization of conditions for the observation of the starch end-point. The variation would be expected to be very much greater with the normal observation in a flask receiving illumination from all directions and observed from above.

V. REFERENCES

- ABEL, E. (1912).—*Z. anorg. Chem.* **74**: 396.
ABEL, E. (1950).—*Mh. Chem.* **81**: 346.
BRAY, W. C. (1910).—*J. Amer. Chem. Soc.* **32**: 932.
BRAY, W. C., and CONNOLLY, E. L. (1911).—*J. Amer. Chem. Soc.* **33**: 1485.
DODD, G., and GRIFFITH, R. O. (1949).—*Trans. Faraday Soc.* **45**: 551.
EVANS, D. P. (1947).—*Analyst* **72**: 99.
EVANS, D. P., and SIMMONS, N. T. (1944).—*J. Soc. Chem. Ind.* **63**: 29.

- FOULK, C. W., and BAWDEN, A. T. (1926).—*J. Amer. Chem. Soc.* **48** : 2045.
- FROST, H. F. (1944).—*Analyst* **69** : 90.
- GILBERT, G. A., and MARRIOTT, J. V. R. (1948).—*Trans. Faraday Soc.* **44** : 84.
- GUZMAN, J. (1935).—*An. Soc. espan. Fis. Quim.* **33** : 109.
- GUZMAN, J., and RANCANO, A. (1934).—*An. Soc. espan. Fis. Quim.* **32** : 590.
- KOLTHOFF, I. M. (1921).—*Z. anal. Chem.* **60** : 328.
- KOLTHOFF, I. M., and LINGANE, J. J. (1941).—"Polarography." p. 448. (Interscience Publishers Inc. : New York.)
- KOLTHOFF, I. M., and SANDELL, E. B. (1936).—"Quantitative Inorganic Analysis." p. 591. (John Wiley & Sons : New York.)
- LA MER, V. K., and TOMLINSON, H. M. (1937).—*Industr. Engng. Chem. (Anal. Ed.)* **9** : 588.
- POLISSAR, M. J. (1935).—*J. Chem. Educ.* **12** : 89.
- RASCHIG, F. (1915).—*Ber. dtsh. chem. Ges.* **48** : 2088.
- ROSIN, J. (1946).—"Reagent Chemicals and Standards." p. 64. (Van Nostrand Inc. : New York.)
- SALOMON, E. (1897a).—*Z. phys. Chem.* **24** : 55.
- SALOMON, E. (1897b).—*Z. Elektrochem.* **4** : 71.
- TOMLINSON, H. M., and CIAPETTA, F. G. (1941).—*Industr. Engng. Chem. (Anal. Ed.)* **13** : 539.
- YOUNG, S. W. (1904).—*J. Amer. Chem. Soc.* **26** : 1028.

THE VISCOSITY OF UREA SOLUTIONS OF KERATIN

By E. F. WOODS*

[*Manuscript received April 30, 1952*]

Summary

The viscosity-concentration relationships for wool dispersed in concentrated urea solutions are comparable with those reported for globular proteins in urea.

The axial ratios calculated from the viscosity increments by means of the Simha equation, using the assumption of unhydrated prolate ellipsoids as molecular models, vary from 9.9 to 16.3, these values depending on the buffer ion present, the conditions of equilibration, and the concentration of urea. There appears to be a slight change in configuration to a more spherical molecule on removal of salts from the urea solutions and on reduction of the urea concentration.

I. INTRODUCTION

Owing to the extreme insolubility of the keratins there have been comparatively few studies on their size and shape in solution. From the results of such studies on wool (Olofsson and Gralén 1947; Speakman 1949; Alexander 1950; Alexander and Earland 1950; Mercer and Olofsson 1951) it is clear that the nature of the products obtained depends on the method used to break the disulphide bonds and also on the extent of peptide bond rupture. Following disulphide bond rupture the possibility of the bond reforming in some other position or a new linkage such as the lanthionine linkage (Horn, Jones, and Ringel 1941) being formed may add to the variety of products.

The measurement of the viscosity may be used as a comparatively simple method of estimating the degree of asymmetry of macromolecules in solution and it has been widely employed for the characterization of linear polymers, which show large viscosity increments in solution. Both native and denatured proteins have also been studied by this method although the viscosity increments are much smaller than those observed with polymers, except those of a few very asymmetric protein molecules such as tobacco mosaic virus and myosin. In the present work the viscosity was investigated of extracts of Merino wool prepared by partial reduction of wool with sodium bisulphite in concentrated urea solution. Mercer and Olofsson (1951) have reported sedimentation studies on solutions prepared in a similar way.

II. EXPERIMENTAL

(a) Preparation of Keratin Dispersion

The wool used was a 64's Merino which had been scoured with ethanol for 7 hr. in a soxhlet extractor and washed in water. For reduction at pH 5.7, 50 g. of wool was extracted with 300 ml. 8M urea (laboratory grade) containing

4 per cent. sodium bisulphite (Merck, A.R.) at 50 °C. for 2 days. The extract, which set to a gel, was diluted by adding more 8M urea and 500 ml. of solution was recovered.

For reduction at higher pH values the method of Mercer and Olofsson (1951) was followed. Twenty g. of wool was cut into short lengths, suspended in 350 ml. saturated urea solution at 20 °C. (9.5M) containing 5 per cent. sodium bisulphite; the pH was adjusted to 8.0 by the addition of solid sodium bicarbonate and sodium carbonate, and the suspension was allowed to stand with occasional shaking for 24 hr. at 50 °C. About 270 ml. of solution was recovered and the final pH was 9.25. In each method of extraction about 30 per cent. of the original wool dissolved.

(b) *Measurement of Viscosity*

Prior to the determination of viscosity, the solutions obtained as described above were dialysed in cellophane tubing against successive daily changes of the urea solution in which the viscosity was to be measured, and finally equilibrated for a period of one week without changing the solution. Protein concentrations were determined as follows: 5 ml. aliquots of the solution were dialysed in a cellophane bag against running water to remove the urea and buffer salts, and the contents weighed after drying at 105–110 °C. for 18 hr.

Relative viscosity measurements were made in capillary viscometers (British Standard viscometers Nos. 0 and 1 were used), the usual precautions for accurate measurements being observed. The temperature of measurement was 25 ± 0.02 °C. The relative viscosity η_{rel} was measured at a number of protein concentrations. The intrinsic viscosity $[\eta]$ was determined by extrapolation to zero concentration of the plot of the reduced viscosity η_{sp}/C against C , where C = concentration in g. 100 ml. and the specific viscosity $\eta_{sp} = \eta_{rel} - 1$. The viscosity increment v , assuming no hydration, was calculated from the equation $v = 100[\eta]/V$, where V is the partial specific volume of dispersed keratin. The latter was assumed to be 0.703, the value reported for wool in sodium sulphide solution (Olofsson and Gralén 1947).

(c) *Measurement of Osmotic Pressure*

The osmotic pressure was measured using the apparatus described by Bull (1941) except that the toluene capillary was attached to the osmometer by means of a ground glass joint. The measurements were made at 25 °C. Before use, the collodion membranes were soaked in the urea-buffer mixture. After each experiment the protein concentration was determined inside and outside the membrane. The molecular weights are only apparent values. It would be necessary to determine the osmotic pressure at several concentrations and extrapolate to zero concentration to obtain the true molecular weights.

III. RESULTS

(a) *Viscosity of Solutions Prepared by Reduction at pH 5.7*

(i) In this series of experiments the solutions were dialysed against 8M urea until free of salts and the relative viscosity was measured at a number of concentrations. Buffer salts were then added to bring the solution and solvent

to the required pH and the viscosity was again determined. Table 1 gives the intrinsic viscosities and also the viscosity increments and axial ratios calculated from them, assuming no hydration. The axial ratios are based on the Simha

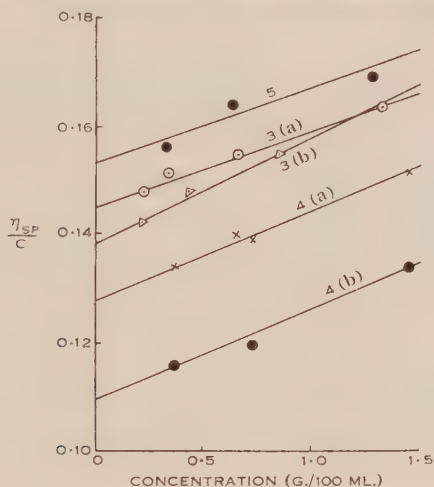


Fig. 1. — Reduced viscosity - concentration relationships for keratin dispersions in urea. Numbers refer to solutions listed in Table 2.

equation (Simha 1940), using the assumption of rigid prolate ellipsoids as molecular models, and were obtained from the tables published by Mehl, Oncley, and Simha (1940).

TABLE 1
VISCOSITY OF SOLUTIONS OF KERATIN PREPARED BY REDUCTION AT pH 5.7 AND DIALYSIS TO FREE FROM SALTS

Soln. No.	Salt or Buffer Added	Urea Concentration (M)	pH	Intrinsic Viscosity $[\eta]$	Viscosity Increment ν	Axial Ratio a/b
1	None	8	7.9	0.103	14.8	10.6
2 (a)	0.1M KCl 0.06M KH_2PO_4 + K_2HPO_4	8	8.1	0.095	13.5	9.9
(b)*	0.1M NaAc + HAc 0.1M KCl	8	5.3	0.099	14.0	10.2

* Prepared from 2 (a) by dialysis of the solution containing phosphate buffer against urea-acetate buffer of pH 5.3.

(ii) In another series of experiments the buffer salts were present during all stages of equilibration, the pH values of the solutions being changed by successive dialysis against urea-buffer mixtures. Figure 1 shows the plot of the

reduced viscosity against concentration for this series of experiments. All the solutions tested revealed a slight linear dependence of the reduced viscosity on concentration. The viscosity increments, calculated from the intrinsic viscosities, and the axial ratios are given in Table 2.

TABLE 2

VISCOSITY OF SOLUTIONS REDUCED AT pH 5.7 AND EQUILIBRATED IN THE PRESENCE OF BUFFER SALTS

Soln. No.	Buffer Present during Dialysis	Urea Concentration (M)	pH	Intrinsic Viscosity $[\eta]$	Viscosity Increment ν	Axial Ratio a/b
3 (a)	0.01M KCl 0.1M NaAc + HAc	8	5.4	0.146	20.7	13.3
(b)*	0.1M KCl 0.067M KH_2PO_4 + K_2HPO_4	8	7.8	0.138	19.6	12.9
4 (a)	0.2M KCl	8	7.3	0.128	18.2	12.2
(b)†	0.2M KCl	5	6.8	0.109	15.5	11.4
5	0.1M KCl 0.067M KH_2PO_4 + K_2HPO_4	8	7.9	0.153	21.7	13.7
6	0.1M NaHSO_3 + Na_2CO_3 + NaHCO_3	5	8.9	0.120	17.1	11.7
7	0.5M NaHSO_3	8	4.8	0.190	27.0	15.8

* Prepared from 3 (a) by successive equilibration against solution of higher pH.

† Prepared from 4 (a) by successive equilibration against solution of lower urea concentration.

From these results it is seen that there is only a slight change in the viscosity increment with variation in the pH of the solution between the limits 4.8 and 8.1. Solutions in which electrolytes were removed during equilibration (Table 1), even though salts were added later, showed smaller viscosity increments than solutions in which electrolyte was maintained during the equilibration (Tables 2 and 3). This irreversible decrease in viscosity increment on removal of electrolyte is not of sufficient magnitude to be attributed to dissociation of the molecules and it is most probably due to a change in configuration of the molecules which can be stabilized by the re-formation of sulphur-containing bonds. When salt was completely removed the solutions in urea became more turbid. The viscosity increment appeared to vary slightly with change in the buffer ions present during the equilibration, acetate buffers giving the highest viscosity

increment of those examined. The results obtained with solutions Nos. 6 and 7 (Table 2) show the effect of maintaining sodium bisulphite in the solution. The reaction of sodium bisulphite on wool can be represented by the equation :



(Middlebrook and Phillips 1942). If the urea-bisulphite extracts of wool are at pH 5, the optimum pH value for this reaction, the re-formation of cross-links should be prevented and it might be expected that the viscosity increment would be lower. However, the values which are reported in Table 2 do not support this expectation ; in fact, at pH 4.8 the viscosity is higher than at pH 8.9.

The reduction of the urea concentration from 8M to 5M lowered the viscosity increment. Similar results have been reported for solutions of water-soluble proteins in urea ; for example, the viscosity increment of serum albumin at various urea concentrations was shown by Neurath and Saum (1939) to increase steadily with increasing concentrations of urea. The reduced viscosity-concentration relationship and the effect of urea concentration on the viscosity resemble those described for globular proteins, although in most proteins studied the reduced viscosity is independent of concentration for solutions containing less than 1 per cent. protein.

(b) *Viscosity of Solutions Prepared by Reduction at pH 8.0*

The preceding results differ markedly from those deduced from sedimentation studies by Mercer and Olofsson (1951). It was thought that this may have been due to reduction at the lower pH value, 5.7, and extracts were therefore prepared at a higher pH value. Table 3 shows the results obtained for the viscosity of solutions prepared by reduction at an initial pH of 8.0. The

TABLE 3
VISCOSITY OF SOLUTIONS REDUCED AT pH 8.0

Soln. No.	Buffer Present during Dialysis	Urea Concentration (M)	pH	Intrinsic Viscosity $[\eta]$	Viscosity Increment ν	Axial Ratio a/b
8	0.1M NaHSO ₃ + Na ₂ CO ₃ + NaHCO ₂	5	9.1	0.149	21.3	13.5
9	0.02M NaAc + HAc	6	6.1	0.200	28.4	16.3

pH of the solution after reduction was 9.25. Although the viscosity increments were higher than those obtained by reduction at an acid pH value they did not approach values which would be expected for the solutions if the molecules were highly asymmetric.

(c) *Osmotic Pressure Measurements*

The osmotic pressure was measured of a keratin solution in 8M urea and 0.2M KCl (solution 4 (a) of Table 2). The pressure reached a maximum on

the first day and dropped slowly for more than a week. It was found that this was due to the slow diffusion of protein through the membrane. The concentration was determined inside and outside the membrane and, from the maximum osmotic pressure obtained by extrapolating the pressure-time curve back to zero time, an apparent molecular weight of 30,000 was calculated. This can only be considered to be approximate. At least one-third of the material diffused through the collodion membrane in the osmometer and it was calculated that this diffusible material would have an average molecular weight of about 10,000, while the non-diffusible fraction would have an average molecular weight of about 50,000. However, the actual values of the molecular weights obtained are not of great significance due to the polydispersity of the keratin dispersion and the fact that the presence of low-molecular weight material prevented the attainment of equilibrium in the osmotic pressure determinations. This latter difficulty was encountered in all osmotic pressure experiments on urea-bisulphite extracts of wool.

IV. DISCUSSION

The relative viscosity-concentration relationships of the keratin dispersions and the magnitude of the values obtained for the viscosity increments are of the same order as those reported for solutions of globular proteins (Neurath and Saum 1939; Bull 1940; Arthur and Saik 1950). If account is taken of hydration, the values for the viscosity increments and axial ratios reported in the present paper are reduced. Assuming a value of 0.4 for the degree of hydration, that is 0.4 g. of water combined with 1 g. of protein, the a/b values are all reduced by approximately 30 per cent. Since the solutions are polydisperse the values obtained for asymmetry do not refer to a single molecular species but will represent a mean value.

The frictional ratios and molecular weights of dispersed wool keratin prepared by reduction of the disulphide bond have been determined from sedimentation and diffusion studies. For fully reduced wool in sodium sulphide Olofsson and Gralén (1947) obtained a molecular weight of 10,000 and a value of 1.77 for f/f_0 corresponding to a value of a/b of 20:1 using the assumption of unhydrated prolate ellipsoids. Mercer and Olofsson (1951) reported values of 3.86 for f/f_0 and 90:1 for a/b for partially reduced wool in concentrated aqueous urea, and they also showed that the solution was polydisperse and contained material of molecular weight 84,000. The results of the present investigation show that urea-bisulphite extracts are also polydisperse but the values of a/b are less than those previously reported for both partially and fully reduced wool. The value 3.86 for f/f_0 obtained by Mercer and Olofsson (1951) is extremely high for proteins. This value is derived from the diffusion coefficient and there is some evidence that the viscosity correction $D_0 - D\eta_0/\eta$, where D and D_0 represent the diffusion coefficients, and η and η_0 the viscosities in urea solution and water respectively, is inapplicable to concentrated urea solutions. Neurath, Cooper, and Erickson (1942) found that frictional ratios when determined from diffusion measurements in urea solutions were higher than those determined from viscosity measurements. More recently Tsao, Bailey, and Adair (1951) have shown that

at extremely low shear rates the viscosity of strong solutions of urea is anomalous, due to structural factors. The values of the frictional ratios of wool dispersions calculated from the diffusion coefficients determined in concentrated urea solution are therefore probably too high. In the present study the velocity gradient in the viscometers used was sufficiently high to overcome structural effects of the solvent and the values obtained by this method should therefore be reliable.

The properties of urea-bisulphite extracts of wool have been described by Mercer (1949). The solutions prepared by reduction at pH 8.0 could be precipitated by adding ammonium sulphate and the curdy precipitates obtained could be drawn into fibres which gave an α -type X-ray photograph. In the present investigation the precipitates obtained by adding ammonium sulphate to the extracts, prepared at both pH 5.7 and 8.0, did not possess fibre-forming properties. It is possible that our derivative of wool corresponds to a non-fibre-forming material obtained by Mercer and Olofsson (1951) at low temperatures and at pH values of 6-7.

The dispersions obtained in this study do not resemble in behaviour other structure-forming proteins, such as tropomyosin or actin. The presence of salts during equilibration of the urea solution exerts a slight influence on the viscosity increment, but this is most likely due to slight changes in the configuration of the molecules, and does not indicate dissociation as is observed with tropomyosin under the influence of electrolytes or concentrated urea (Tsao, Bailey, and Adair 1951). There is also no suggestion of an equilibrium between globular and fibrous molecules as in urea solutions of actin (Szent-Gyorgi and Joseph 1951).

By oxidation of wool with peracetic acid and extraction with ammonia, Alexander (1950) found that 90 per cent. of the wool dissolved and of this soluble portion 70 per cent. was of molecular weight 70,000 and consisted of almost spherical molecules. The remaining 30 per cent. was of molecular weight 4500. A comparison between molecular weights and shapes determined in aqueous solutions prepared in this manner and in strong urea solutions cannot justifiably be made. Urea can have a variety of effects on proteins causing dissociation, unfolding, or aggregation (Neurath *et al.* 1944), and gelation in concentrated protein solutions (Huggins, Tapley, and Jensen 1951). It is interesting to note that wool, when extracted with concentrated urea-sodium bisulphite, formed a solution which gelled at 50 °C. but was dissolved by dilution with more urea solution. Middlebrook (1951) recently determined the chain weight of wool keratin by quantitative estimation of the amino-acid end-groups. He obtained a molecular weight of 60,000 and proposed a molecular model in accordance with these values. He considers that treatment with urea-bisulphite splits the wool into chains of 30,000 but does not suggest that a further splitting into smaller units of 10,000 takes place in the presence of urea-bisulphite. The results reported in this investigation and the results of Mercer and Olofsson (1951), however, show that both high and low molecular weight material is obtained by the action of urea-bisulphite on wool. The results of all the previously reported methods of determining the chain weight of wool keratin give

a material of molecular weight between the limits 60,000–80,000, together with smaller units of molecular weight 4500–10,000. The results reported in the present paper are in fair agreement with these values.

V. REFERENCES

- ALEXANDER, P. (1950).—*Proc. R. Soc. Med.* **44**: 389.
ALEXANDER, P., and EARLAND, C. (1950).—*Nature* **166**: 396.
ARTHUR, J. C., and SAIK, B. H. (1950).—*J. Colloid Sci.* **5**: 326.
BULL, H. B. (1940).—*J. Biol. Chem.* **133**: 39.
BULL, H. B. (1941).—*J. Biol. Chem.* **137**: 143.
HORN, N. H., JONES, W. B., and RINGEL, S. J. (1941).—*J. Biol. Chem.* **138**: 141.
HUGGINS, C., TAPLEY, D. F., and JENSEN, E. V. (1951).—*Nature* **167**: 592.
MERCER, E. H. (1949).—*Nature* **163**: 18.
MERCER, E. H., and OLOFSSON, B. (1951).—*J. Polymer Sci.* **6**: 671.
MEHL, J. W., ONCLEY, J. L., and SIMHA, R. (1940).—*Science* **92**: 132.
MIDDLEBROOK, W. R. (1951).—*Biochem. Biophys. Acta* **7**: 547.
MIDDLEBROOK, W. R., and PHILLIPS, H. (1942).—*Biochem. J.* **36**: 428.
NEURATH, H., and SAUM, A. M. (1939).—*J. Biol. Chem.* **128**: 359.
NEURATH, H., COOPER, G. A., and ERICKSON, J. O. (1942).—*J. Biol. Chem.* **142**: 249, 265.
NEURATH, H., GREENSTEIN, J. P., PUTNAM, F. W., and ERICKSON, J. O. (1944).—*Chem. Rev.* **34**: 157.
OLOFSSON, B., and GRALÉN, N. (1947).—*Proc. 11th Int. Conf. Chem. London*.
SIMHA, R. (1940).—*J. Phys. Chem.* **44**: 25.
SPEAKMAN, J. B. (1949).—"Fibre Science," p. 282. (Textile Institute: Manchester.)
SZENT-GYORGI, A. G., and JOSEPH, R. (1951).—*Arch. Biochem. Biophys.* **31**: 90.
TSAO, T. C., BAILEY, K., and ADAIR, G. S. (1951).—*Biochem. J.* **49**: 27.

ALKALOIDS OF THE AUSTRALIAN RUTACEAE: *PENTACERAS AUSTRALIS* HOOK. F.

II. IDENTIFICATION OF 5-METHOXYCANTHINONE

By EVA R. NELSON* and J. R. PRICE†

[Manuscript received March 18, 1952]

Summary

The alkaloid $C_{15}H_{10}O_2N_2$, present in the bark of *Pentaceras australis* Hook. f., has been shown to be 5-methoxycanthin-6-one. It is oxidized by potassium permanganate to β -carboline-1-carboxylic acid and its lactam ring may be opened by alkali to give a β -carbolylmethoxyacrylic acid which readily recyclizes to the alkaloid. The position of the methoxyl group is demonstrated by demethylation and condensation of the resulting hydroxy-compound with *o*-phenylenediamine to give a hydroxyquinoxaline. Attempted hydrogenation of the alkaloid or the derived methoxyacrylic acid led to elimination of methoxyl.

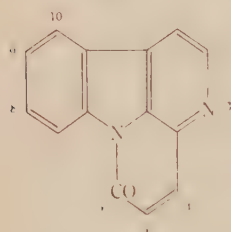
I. INTRODUCTION

Pentaceras australis Hook. f. was shown by Haynes, Nelson, and Price (1952) to contain three alkaloids, $C_{14}H_8ON_2$, $C_{15}H_{10}O_2N_2$, and $C_{15}H_{10}OSN_2$. The first, which they showed to be canthin-6-one (I), provided an example of a new ring system among the plant alkaloids. The present paper establishes the structure of the second alkaloid, $C_{15}H_{10}O_2N_2$. Like canthinone it is an optically inactive, monoacid base forming a *picrate*, $C_{15}H_{10}O_2N_2 \cdot C_6H_3O_7N_3$ and a *methiodide*, $C_{15}H_{10}O_2N_2 \cdot CH_3I$. It contains one methoxyl group but no methylimino-group and cannot be acetylated. Oxidation with permanganate in acetone gives β -carboline-1-carboxylic acid, $C_{12}H_8O_2N_2$, identified by comparison of the methyl ester with a synthetic specimen. Hydrolysis with alcoholic potash gives a hygroscopic yellow acid from which the alkaloid is readily regenerated. It follows that the alkaloid is a methoxycanthinone with the methoxyl group substituted in the lactam ring at either position 4 or 5. The *cis*-acid resulting from opening the lactam ring is not converted to the corresponding *trans*-acid under the conditions which bring about isomerization of β -carbolylacrylic acid itself (cf. Haynes, Nelson, and Price 1952). The alkaloid is unattacked by ethanolic hydrochloric acid but is demethylated by hydrobromic acid in acetic acid. The deep orange demethylation product, $C_{14}H_8O_2N_2$, is weakly acidic, gives a dull yellowish green colour with ferric chloride, and forms an *acetyl derivative*, $C_{14}H_7ON_2(OOC.CH_3)$, insoluble in cold aqueous sodium hydroxide.

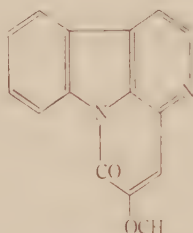
* Chemistry Department, University of Melbourne; present address: Nicholas Pty. Ltd., Melbourne.

† Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

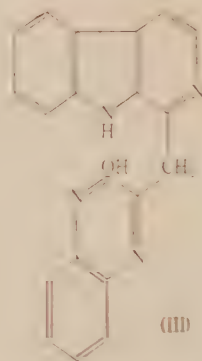
The alkaloid is regenerated by methylation of the acetyl derivative with diazomethane in moist ether. Though methoxycanthinone does not react with *o*-phenylenediamine, its demethylation product does, forming an orange coloured condensation product $C_{20}H_{14}ON_4$, thus establishing that the hydroxyl group, and therefore the methoxyl group in the alkaloid, is situated in the *ortho*-position



(I)



(II)



(III)

with respect to the $-CO-$ group: that is, the alkaloid is 5-methoxycanthin-6-one (II), and the acid resulting from opening the lactam ring is 1-methoxy-2-(1- β -carboly)acrylic acid. The ultraviolet absorption spectra of the alkaloid and of 5-acetoxycanthinone (see Fig. 1 and Table 1) closely resemble one another and that of canthinone itself. The molecular formula of the condensation product

TABLE I
ULTRAVIOLET ABSORPTION DATA

Substance	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$\log \epsilon_{max}$
5-Methoxycanthin-6-one	372	4.35	~ 298	3.9	~ 259	4.0
	354.5	4.29	290	3.96	246.5	4.42
	339.5	3.99			~ 243	4.38
5-Acetoxycanthin-6-one	377	4.16	~ 300	3.93	~ 267	3.98
	359	4.21	294	3.94	~ 257	4.07
	345	3.99				
Canthin-6-one	381	4.14	299	3.91	269	4.03
	362	4.17	~ 294	3.90	259	4.05
	347	3.94			251	4.09

of 5-hydroxycanthinone and *o*-phenylenediamine indicates that it is a 2-hydroxy-quinoxaline of structure III and not a phenazine derivative. Its formation parallels that of 2-hydroxy-3-(*o*-aminophenyl)quinoxaline from isatin reported by Marchlewski (1899) and by Marchlewski and Sosnowski (1901). An attempt was made to synthesize 5-hydroxycanthinone by condensing harman with oxalic

ester in the presence of potassium ethoxide (see Kleisinger and Wislicenus 1909) but no reaction occurred. Leonard and Boyer (1950) likewise failed to condense 5-methylacridine with oxalic ester.

Hydrogenation of canthinone and β -carbolylacrylic acid took place smoothly with the formation of 4,5-dihydrocanthinone and β -carbolypropionic acid respectively. 5-Methoxycanthinone, however, and the β -carbolylmethoxyacrylic acid derived from it, behaved differently. Hydrogenation of the alkaloid over Raney nickel was not successful; it was either recovered unchanged or,

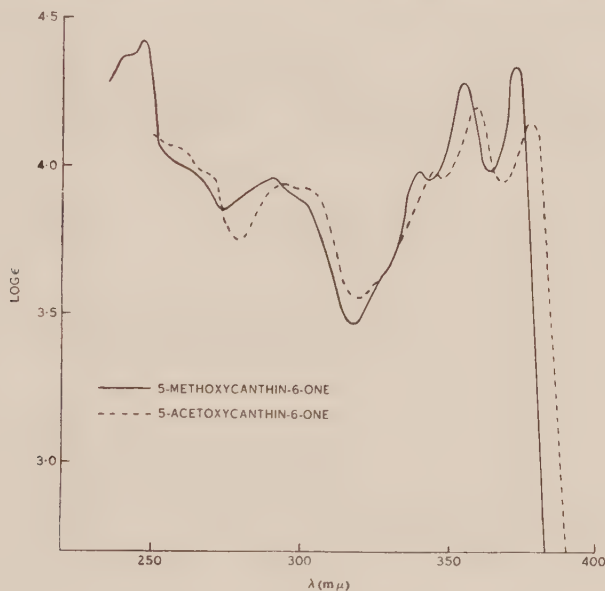


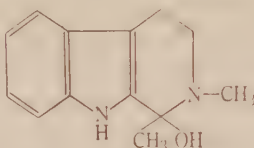
Fig. 1

when a benzene solution was refluxed with excess Raney nickel, the bulk of the material was lost, presumably through retention of the reaction product by the nickel. Reduction with zinc and acetic acid gives 4,5-dihydrocanthinone, but if the reaction time is short, some canthinone also is obtained. This unexpected elimination of methoxyl is likewise encountered with β -carbolylmethoxyacrylic acid which with Raney alloy and sodium hydroxide gives β -carbolypropionic acid. However, reduction of 5-hydroxycanthinone with Raney alloy and sodium hydroxide follows a normal course giving the anticipated β -carbolyllactic acid.

Canthinone and 5-methoxycanthinone methiodides are both attacked by hot alkali and in the latter case the yellow product has been isolated and found to possess the molecular formula $C_{13}H_{14}ON_2$. The substance contains a methylimino-group but no methoxyl and the carbinol-base structure IV is suggested for it. Such a structure could arise by fission of the C_1-N^* bond followed by

* The carbon atom referred to is C_1 of the β -carboline nucleus, an atom not numbered in the canthinone system (I).

demethylation of the methoxyl group and elimination of two carbon atoms from the resulting 1,3-diketo-acid.



(IV)

Pharmacological tests carried out by Associate Professor F. H. Shaw showed 5-methoxycanthinone to have no notable physiological effect on mice or toads. 5-Methoxycanthinone methiodide exhibited weak curariform activity.

II. EXPERIMENTAL

All melting points are corrected except where otherwise stated. Microanalyses were carried out by the C.S.I.R.O. Microanalytical Laboratory.

(a) Properties of the Alkaloid $C_{15}H_{10}O_2N_2$

The alkaloid is easily soluble in chloroform, less soluble in benzene, and only sparingly soluble in cold ethanol or acetone. It is conveniently crystallized from chloroform-ethanol.

Found: CH_3O , 12.1%; CH_3N , nil.

Calculated for $C_{15}H_{10}O_2N_2$: CH_3O , 12.4% (one methoxyl).

A solution in chloroform (c. 1.025) was optically inactive. The alkaloid is only sparingly soluble in 0.01N HCl, but dissolves in 0.1N HCl giving a pale yellow solution. The hydrochloride crystallized from 1N HCl as fine yellow needles (not analysed) which melted at 206–207°C. (uncorr., decomp.) when the temperature was raised rapidly; it dissolves in water but the free base separates on standing. The *picrate*, yellow needles from methanol, melted at 242–244°C. after sintering at 235°C.

Found: C, 52.7; H, 2.8; N, 14.5%.

Calculated for $C_{15}H_{10}O_2N_2 \cdot C_6H_5O_7N_3$: C, 52.6; H, 2.7; N, 14.6%.

The *methiodide*, prepared by refluxing a solution of the alkaloid (0.1 g.) and methyl iodide (5 ml.) in chloroform (25 ml.), crystallized from water as orange-red flat needles, m.p. 308–309°C. (uncorr., decomp.).

Found: C, 49.3; H, 3.3; I, 32.1; CH_3O , 7.9; CH_3N , 7.3%.

Calculated for $C_{16}H_{12}O_2N_2I$: C, 49.0; H, 3.3; I, 32.4; CH_3O , 7.9 (one methoxyl); CH_3N , 7.4%.

(b) Oxidation with Permanganate

5-Methoxycanthinone was oxidized slowly at room temperature by permanganate in acetone, the reaction being completed after c. 18 hours. The precipitated manganese dioxide was suspended in water and dissolved with sulphur dioxide leaving a yellow acid which was purified by precipitating the colourless sodium salt from 10% sodium hydroxide and finally by crystallization from acetic acid. It was obtained as yellow needles, m.p. 239.5°C. (decomp.).

Found: C, 67.8; H, 3.7; N, 13.1%; CH_3O , nil.

Calculated for $C_{12}H_8O_2N_2$: C, 67.9; H, 3.8; N, 13.2%.

Esterification by methanolic hydrogen chloride gave the methyl ester, colourless needles from aqueous methanol, m.p. 167–167.5°C. undepressed by admixture with a synthetic specimen of the methyl ester of β -carboline-4-carboxylic acid. Decarboxylation gave β -carboline, m.p. and mixed m.p. 200–201°C.

(c) *The Action of Alkali on 5-Methoxycanthinone*

The alkaloid is not soluble in cold aqueous sodium hydroxide but dissolves slowly on heating giving a yellow solution. 1-Methoxy-2-(1'- β -carbonyl)acrylic acid was prepared by adding the alkaloid (1 g.) to a boiling solution of potassium hydroxide (10 g.) in ethanol (100 ml.), distilling off the bulk of the solvent, and completing its removal azeotropically with water. The solution was acidified with acetic acid and the precipitated acid crystallized from acetic acid as a micro-crystalline hygroscopic powder, m.p. 235 °C. (decomp.).

Found: C, 59.6; H, 5.1%.

Calculated for $C_{11}H_{12}O_3N_2 \cdot 2H_2O$: C, 59.2; H, 5.3%.

Heating an aqueous solution of the sodium salt of this acid resulted in the crystallization of 5-methoxycanthinone, m.p. and mixed m.p. 241.5–242 °C. The alkaloid was also regenerated from the acid by boiling an acetic acid solution for 3 hours and after attempted esterification with methanolic hydrogen chloride. The acid was evidently not isomerized by boiling with 10% sodium hydroxide for 7 hours, followed by evaporation to c. $\frac{1}{4}$ the volume, as 5-methoxycanthinone (m.p. and mixed m.p.) was recovered from the sodium salt which separated on cooling.

(d) *5-Hydroxycanthinone*

(i) 5-Methoxycanthinone (2.5 g.), aqueous hydrobromic acid (45%; 25 ml.), and acetic acid (50 ml.) were refluxed for 24 hours. After cooling, the solid, which had partly separated from the hot solution, was filtered off, triturated with water, and crystallized from acetic acid giving golden yellow needles of what was evidently a salt. This was hydrolysed by titration with water to a deep orange crystalline powder, m.p. 259–261 °C. (decomp.). Yield 2.0 g.

Found: C, 71.4; H, 3.5; N, 11.7%; CH_3O , nil.

Calculated for $C_{14}H_8O_2N_2$: C, 71.2; H, 3.4; N, 11.8%.

5-Hydroxycanthinone gives a dull yellowish green colour with ferric chloride. It is only slightly soluble in aqueous sodium bicarbonate, but dissolves in sodium carbonate forming a yellow sodium salt sparingly soluble in the presence of excess alkali.

(ii) 5-Hydroxycanthinone (1 g.) was refluxed with acetic anhydride* (5 ml.) for 2 hours. 5-Acetoxyacanthinone separated on cooling and was obtained as colourless needles, m.p. 231.5–232 °C., after recrystallization from benzene. Yield 0.67 g.

Found: C, 69.2; H, 3.5; N, 10.1; CH_3CO , 15.7%.

Calculated for $C_{14}H_7ON_2(OOC.CH_3)$: C, 69.1; H, 3.6; N, 10.1; CH_3CO , 15.5%.

On standing, the substance slowly becomes yellow. It is insoluble in cold 2% sodium hydroxide.

(iii) 5-Acetoxyacanthinone (0.2 g.) in dioxan (20 ml.) was allowed to stand for 5 days with a solution of diazomethane (from 0.7 g. N-nitrosomethylurea) in moist ether (20 ml.), the orange solution evaporated to dryness under reduced pressure, and the residue treated with 5 ml. 5% sodium hydroxide solution. Extraction with chloroform and crystallization from ethanol gave 5-methoxycanthinone, m.p. 241–242 °C. alone or mixed with the alkaloid. The yield was almost quantitative. The identity was confirmed by preparation of the picrate, m.p. and mixed m.p. 241–242 °C.

(e) *Reaction of 5-Hydroxycanthinone with o-Phenylenediamine*

A solution of 5-hydroxycanthinone (0.5 g.) and o-phenylenediamine (0.25 g.) in acetic acid (12 ml.) was refluxed for $\frac{1}{2}$ hour, cooled, and filtered from unchanged hydroxycanthinone (0.18 g.). The deep red solution was poured into water, neutralized with sodium bicarbonate, and the mixture shaken with chloroform. The solid which had separated was filtered off, washed with chloroform and water, and the filtrate and washings discarded. The dried solid, dissolved in a small volume of pyridine, was run through a short column of alumina, the eluate evaporated to dryness in a vacuum at room temperature, and the resulting reddish orange needles triturated

* If pyridine is added to catalyse the acetylation, the reaction mixture rapidly turns black.

with acetone and filtered. Yield 0.1 g. The substance, which is sparingly soluble in most organic solvents but very soluble in pyridine and in acetic acid, melted at 343–345 °C. (uncorr., decomp.).

Found: C, 73.4; H, 4.4; N, 17.1%.

Calculated for $C_{20}H_{14}ON_4$: C, 73.6; H, 4.3; N, 17.2%.

(f) *Hydrogenation of 5-Methoxycanthinone*

(i) 5-Methoxycanthinone was recovered unchanged (m.p. and mixed m.p.) after attempted hydrogenation in ethanolic solution at room temperature in the presence of Raney nickel.

(ii) A solution of the alkaloid (0.5 g.) in benzene (50 ml.) was refluxed for 2 hours with Raney nickel (c. 4 g.), decanted, and the nickel washed several times by decantation with benzene. Evaporation of the combined benzene solution and washings gave 0.09 g. colourless solid, identified by m.p. and mixed m.p. as unchanged 5-methoxycanthinone.

(iii) 5-Methoxycanthinone (2.4 g.) in acetic acid (50 ml.) was boiled for 15 minutes with granulated zinc, the reaction mixture poured into water, basified with sodium bicarbonate, and extracted with chloroform. The residue (2.33 g.), after evaporation of the chloroform, was dissolved in the minimum amount of benzene and treated with light petroleum giving a precipitate consisting essentially of unchanged alkaloid. After removal of the solvent, the light petroleum-soluble material (1.20 g.) was dissolved in benzene and chromatographed on alumina, 10 fractions being taken as follows:

Fraction	Weight (g.)	M.p. (°C.)
1	0.15	120–124.5
2	0.44	122–124
3	0.09	100–102
4	0.06	129–142
5	0.08	143–154
6	0.10	149–152
7	0.07	138–147
8	0.03	141–142
9	0.02	
10	0.01	

Fractions 1 and 2 crystallized from light petroleum as colourless needles, m.p. 129–130 °C. undepressed by admixture with 4,5-dihydrocanthinone.

Found: C, 75.8; H, 4.5%.

Calculated for $C_{14}H_{10}ON_2$: C, 75.7; H, 4.5%.

Fractions 5, 6, and 7 were combined and the base purified by passage through the picrate and by crystallization from aqueous methanol. It was obtained as colourless needles, m.p. 160–161.5 °C. undepressed by admixture with canthinone.

Found: C, 76.4; H, 3.6%.

Calculated for $C_{14}H_8ON_2$: C, 76.4; H, 3.6%.

When 5-methoxycanthinone in acetic acid was boiled with granulated zinc for 45 minutes and the reaction mixture worked up as above, only dihydrocanthinone was isolated.

(iv) A solution of 5-methoxycanthinone (0.75 g.) in methanol (25 ml.) containing potassium hydroxide (10 g.) was diluted with hot water (100 ml.) and heated on the water bath. Raney alloy (3 g.) was added gradually during 3 hours and the solution filtered and evaporated to a volume of c. 30 ml. The solid which separated on cooling was filtered, dissolved in water, and reprecipitated by the addition of sufficient potassium hydroxide solution to bring the concentration to c. 25%. The filtered salt was redissolved in water, the solution acidified with acetic acid, and the precipitated acid crystallized from aqueous acetic acid. Yield 0.35 g. Esterification with methanolic hydrogen chloride gave the methyl ester of 5-carbolypropionic acid, m.p. 144–145 °C. undepressed by admixture with an authentic specimen. Saponification of the methyl ester gave β -carbolypropionic acid, m.p. 215–216 °C. (decomp.).

(g) *1-Hydroxy-2-(1'- β -carbonyl)acrylic Acid*

A suspension of 5-hydroxycanthinone (0.5 g.) in sodium hydroxide (7.5%; 40 ml.) was heated on the water-bath and reduced by the gradual addition of Raney alloy (2.5 g.) during 5 hours. The solution was then filtered and concentrated to c. 10 ml. The sodium salt which separated on cooling was filtered off, dissolved in water, and the solution acidified with acetic acid. The precipitated acid crystallized from 25% aqueous acetic acid as cream needles, m.p. 200–200.5 °C. (decomp.). Yield 0.22 g.

Found: C, 61.6; H, 5.1%.

Calculated for $C_{14}H_{12}O_3N_2 \cdot H_2O$: C, 61.3; H, 5.1%.

The acid gave no colour with ferric chloride in ethanolic solution. Esterification with methanolic hydrogen chloride gave the methyl ester, almost colourless needles from benzene, m.p. 114–115 °C. Yield 0.48 g. from 1.1 g. acid.

Found: C, 67.1; H, 5.3%.

Calculated for $C_{15}H_{14}O_3N_2$: C, 66.7; H, 5.2%.

(h) *The Action of Alkali on 5-Methoxycanthinone Methiodide*

When 5-methoxycanthinone methiodide was dissolved in 10% aqueous sodium hydroxide and allowed to stand at room temperature no visible change occurred, but when the solution was heated it deposited a mass of yellow crystals. The substance crystallized from water as bright yellow needles, m.p. 190–193 °C. (decomp.).

Found: C, 73.4; H, 6.3; N, 13.1; CH_3N , 13.4%; CH_3O , nil.

Calculated for $C_{13}H_{14}ON_2$: C, 72.9; H, 6.5; N, 13.1; CH_3N , 13.6%.

The substance may be crystallized from benzene. It does not react with 2,4-dinitrophenylhydrazine.

(i) *Attempted Synthesis of 5-Hydroxycanthinone*

Potassium (0.2 g.) under absolute ether (3 ml.) was dissolved by the addition of absolute ethanol (0.6 ml.) and to the mixture was added a solution of ethyl oxalate (0.9 ml.) in absolute ether (2 ml.) followed, dropwise, by a solution of harman (0.9 g.) in anisole (20 ml.). The solution turned yellow and, after standing several days, solid separated. This was filtered off and a second crop obtained by concentration of the filtrate under reduced pressure. Both crops were refluxed for 4 hours with methanolic hydrogen chloride, but yielded only harman, m.p. and mixed m.p. 232–234 °C.

(j) *Ultraviolet Absorption Spectra*

The ultraviolet absorption spectra were measured in dioxan solution by means of a Beckmann model DU spectrophotometer.

III. ACKNOWLEDGMENTS

The authors are indebted to Professor E. J. Hartung for providing accommodation and facilities in the Chemistry Department, University of Melbourne, and to Associate Professor F. H. Shaw, Department of Physiology, University of Melbourne, for carrying out pharmacological tests.

IV. REFERENCES

- HAYNES, H. F., NELSON, EVA R., and PRICE, J. R. (1952).—*Aust. J. Sci. Res. A* **5**: 387.
 KLEISINGER, E., and WISLICENUS, W. (1909).—*Ber. dtsh. chem. Ges.* **42**: 1140.
 LEONARD, N. J., and BOYER, J. H. (1950).—*J. Amer. Chem. Soc.* **72**: 2980.
 MARCHLEWSKI, L. (1899).—*J. prakt. Chem.* [2], **60**: 407.
 MARCHLEWSKI, L., and SOSNOWSKI, J. (1901).—*Ber. dtsh. chem. Ges.* **34**: 1108.

SHORT COMMUNICATIONS

ANGULAR DISTRIBUTION OF α -PARTICLES FROM THE $\text{Li}^7(p,\alpha)\text{He}^4$ REACTION AT LOW PROTON BOMBARDING ENERGIES*

By F. HIRST†

The work on the angular distribution of α -particles from the $\text{Li}^7(p,\alpha)\text{He}^4$ reaction performed by Martin *et al.* (1949) and Hirst and Uebergang (1951) has been extended in the proton energy range 60–250 keV. to determine where the coefficient $A(E)$ arising in the angular distribution expression

$$Y(\theta) = Y(90^\circ)[1 + A(E) \cos^2 \theta + B(E) \cos^4 \theta]$$

vanishes or becomes negative.

This expression implies that both p and f protons take part in the $\text{Li}^7(p,\alpha)\text{He}^4$ reaction and on this basis Inglis (1948) has derived a theoretical expression for $A(E)$ which requires that $A(E)$ becomes zero at two values of the proton energy. For p protons alone, the $B(E)$ term is not present and the theoretical value of $A(E)$ vanishes only once.

In the present investigation it was found that the $A(E)$ curve crossed the abscissa near 100 keV. and the existence of this low energy node together with that discovered by Heydenburg *et al.* (1948a, 1948b) at a higher proton energy (viz. 2.7 MeV.) confirms the participation of f protons in the $\text{Li}^7(p,\alpha)\text{He}^4$ reaction.

The measurements were obtained using nuclear emulsion photographic plates, the camera being the same as used by Martin *et al.* except that a movable target was provided. This allowed the target area to be changed five times during the plate exposure, thereby avoiding excessive deposition of carbon on the target surface. At least five exposures were made at each bombarding energy and, in particular, at 60 keV., 17 photographic plates were needed to record sufficient α -particles. For proton energies between 100 and 250 keV., thin targets (~ 25 keV. in effective thickness) were used, whilst thick targets (~ 60 keV. in effective thickness) were employed for the 60 keV. exposures. Although the thick target result should be correlated with an energy less than the nominal bombarding energy, owing to the uncertainty of the correlation, this result has been displayed at 60 keV. in Figure 1.

As previous experimental evidence (Hirst and Uebergang 1951; Martin *et al.* 1949; Talbott, Busala, and Weiffenbach 1951) had shown that the $B(E)$ term arising in the above angular distribution expression was essentially zero in the proton energy range 60–250 keV., the data were analysed by a least

* Manuscript received March 25, 1952.

† Physics Department, University of Melbourne.

squares method for the $A(E)$ coefficient only. Figure 1 compares the results of this experiment with those of other workers. Observations below a proton energy of 250 keV. are difficult because of the low reaction cross section and previously only Young, Ellett, and Plain (1940) have published extensive results in this energy range. Since 100 keV. was the lowest proton energy at which measurements were recorded, Young, Ellett, and Plain failed to detect the point where $A(E)$ became negative. Figure 1 shows that the present results are consistent with those of Martin *et al.* and of Hirst and Uebergang. In general the data of the latter two investigations were analysed for both $A(E)$ and $B(E)$ coefficients, which accounts for their large statistical uncertainty.

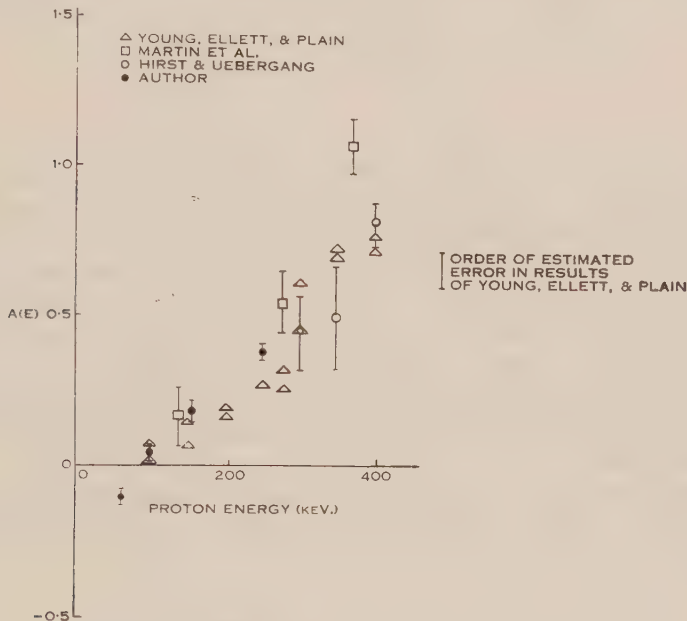


Fig. 1.— $A(E)$ against proton energy.

Although this experiment has defined the $A(E)$ curve more precisely at low proton energies, the assumptions concerning barrier penetrability and the levels of the Be^8 compound nucleus made by Inglis still provide sufficient freedom to account for the experimental observations now available.

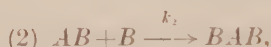
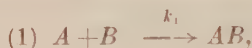
References

- HEYDENBURG, N. P., HUDSON, C. M., INGLIS, D. R., and WHITEHEAD, W. D. (1948a).—*Phys. Rev.* **73**: 241.
 HEYDENBURG, N. P., HUDSON, C. M., INGLIS, D. R., and WHITEHEAD, W. D. (1948b).—*Phys. Rev.* **74**: 405.
 HIRST, F., and UEBERGANG, R. G. (1951).—*Aust. J. Sci. Res. A* **4**: 284.
 INGLIS, D. R. (1948).—*Phys. Rev.* **74**: 21.
 MARTIN, L. H., BOWER, J. C., DUNBAR, D. N. F., and HIRST, F. (1949).—*Aust. J. Sci. Res. A* **2**: 25; *Nature* **164**: 310.
 TALBOTT, F. L., BUSALA, A., and WIEFFENBACH, G. C. (1951).—*Phys. Rev.* **82**: 1.
 YOUNG, V. J., ELLETT, A., and PLAIN, G. J. (1940).—*Phys. Rev.* **58**: 498.

KINETIC TREATMENT APPROPRIATE TO RAPID REACTIONS INVOLVING TWO CONSECUTIVE SECOND ORDER STEPS*

By H. G. HIGGINS† and E. J. WILLIAMS‡

Reactions involving consecutive second order stages have not received much attention from the kinetic standpoint, owing to the complex nature of the integrals obtained from the simultaneous differential equations expressing the reaction rates in general terms. The type of reaction in question is one which may be considered as taking place irreversibly in consecutive stages :



where the reactants A and B are both capable of unifunctional behaviour, and A is also capable of bifunctional behaviour, and where k_1 and k_2 are the velocity constants. It is desirable to be able to express the final concentrations of reactants and products in terms of the initial concentrations and the velocity constants of the separate stages. It will be shown that by eliminating time from the equations and inserting appropriate boundary conditions for the final state in the solutions, relations can be established between the final concentrations and the ratio of the velocity constants. Thus the relative velocities of the two stages can be determined from experimental observation of the initial and final concentrations.

The necessity for the solution of this problem arose in the first instance in attempting to interpret spectrophotometric data on the coupled products of amino acids and diazonium salts.

Theory

(a) *General*.—Let c with appropriate subscript denote the molar concentrations at time t , and x_1, x_2 denote the fall in the concentrations of A and AB respectively in time t due to the individual reactions 1 and 2. The initial concentrations of A and B are a and b respectively, and of AB and BAB are zero.

Then

$$x_1 = c_{AB} + c_{BAB}, \quad \dots \dots \dots (1)$$

and

$$x_2 = c_{BAB}, \quad \dots \dots \dots (2)$$

whence

$$x_1 - x_2 = c_{AB}, \quad \dots \dots \dots (3)$$

* Manuscript received April 10, 1952.

† Division of Forest Products, C.S.I.R.O., Melbourne.

‡ Section of Mathematical Statistics, C.S.I.R.O., at Division of Forest Products, C.S.I.R.O., Melbourne.

Further

$$a - x_1 = c_A \text{ and } b - x_1 - x_2 = c_B. \quad \dots\dots\dots (4)$$

From the law of mass action

$$dx_1/dt = k_1(a - x_1)(b - x_1 - x_2), \quad \dots\dots\dots (5)$$

and

$$dx_2/dt = k_2(x_1 - x_2)(b - x_1 - x_2). \quad \dots\dots\dots (6)$$

Since the quantities produced by the reaction depend on the ratio $k_1 : k_2$ rather than the absolute velocity constants, we may, without loss of generality, put $k_1 = 1$ and $k_2 = K$.

We now put

$$y_1 = a - x_1; \quad y_2 = a - x_2; \quad p = b - 2a,$$

so that

$$c_{AB} = x_1 - x_2 = y_2 - y_1; \quad c_{BAB} = x_2 = a - y_2.$$

Equations (5) and (6) now become

$$dy_1/dt = -y_1(p + y_1 + y_2), \quad \dots\dots\dots (7)$$

$$dy_2/dt = K(y_1 - y_2)(p + y_1 + y_2).$$

On eliminating t , we have

$$\frac{dy_2}{dy_1} = \frac{K(y_2 - y_1)}{y_1}.$$

The initial concentrations of A , B , AB , and BAB are a , b , 0 , and 0 respectively, and since $t = 0$, $x_1 = 0 = x_2$.

When $K \neq 1$, the solution satisfying the initial conditions ($y_1 = a$, $y_2 = a$) is

$$y_2 = \frac{y_1 [K - (y_1/a)^{K-1}]}{K - 1}, \quad \dots\dots\dots (8)$$

and when $K = 1$, it is

$$y_2 = y_1 [1 - \log (y_1/a)]. \quad \dots\dots\dots (9)$$

Substituting for y_2 , from (8), in equation (7), we have

$$-y_1 \{ p + y_1 + y_1 [K - (y_1/a)^{K-1}] / (K - 1) \}.$$

This cannot be solved readily except for special values of K , and even then the results obtained are too complicated to be of much practical use.

(b) *Final State*.—The final concentrations, C , may be derived very simply when either $k_1 = 0$ or $k_1 \neq 0$; $k_2 = 0$, and the solutions when $a > b$ and $a < b$ are included in Figure 1.

The more general case is, however, when $k_1, k_2 \neq 0$ ($K \neq 0, \infty$), which has also to be considered for two relationships between a and b .

(i) When $a < b/2$: Here the initial concentration of B is sufficient for each molecule of AB formed to be finally converted to BAB , so that $x_2 = x_1 = a$, and

$$C_A = C_{AB} = 0; \quad C_B = b - 2a; \quad C_{BAB} = a. \quad \dots\dots\dots (10)$$

(ii) When $a = b/2$: Here the final concentrations of AB and BAB will depend on the *ratio* of the velocity constants, since B must react with A to produce AB , then with AB to produce BAB . As A is present in sufficient concentration initially for all the molecules of B to react eventually, the substances present finally will be A , AB , and BAB . Then

$$x_1 + x_2 = b, \quad \dots\dots\dots (11)$$

and $C_B = 0$.

From (11),

$$y_1 + y_2 = 2a - b = -p; \text{ and } y_2 = -(p + y_1).$$

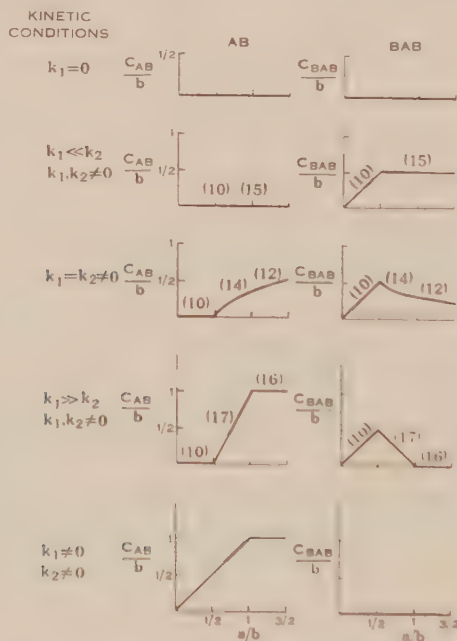


Fig. 1.—Dependence of final concentration of products AB , BAB on initial concentration of A for various kinetic conditions. Numbers shown on curves refer to equations given in the text.

Substituting for y_2 in (8) and (9), we have for $K \neq 1$

$$p + y_1 = -y_1 [K - (y_1/a)^{K-1}] / (K-1), \quad \dots\dots\dots (12)$$

and for $K=1$

$$p + y_1 = -y_1 [1 - \log (y_1/a)].$$

If a and b are given, y_1 can thus be determined as a function of K . Hence if y_1 be observed, K can be determined.

From (4), $C_A = y_1$.

From (3) and (11), $C_{AB} = 2x_1 - b = 2(a - y_1) - b = -(p + 2y_1)$. .. (13)

From (2) and (11), $C_{BAB} = b - x_1 = b - (a - y_1) = p + a + y_1$ (14)

Conforming still to the boundary conditions of the above paragraph, we may consider two special cases, namely, when K is very large (but not ∞) and when K is very small (but not zero), i.e. when $k_2 \gg k_1$ and when $k_1 \gg k_2$ respectively.

$k_2 \gg k_1$: Since stage 2 of the reaction takes place very much faster than stage 1, a molecule of B may be considered to react with AB , as soon as this is formed, to produce BAB , in preference to reacting with A to form another molecule of AB . Thus

$$C_{AB}=0; \quad C_{BAB}=b/2; \quad C_A=a-b/2. \quad \dots\dots\dots (15)$$

$k_1 \gg k_2$: Since stage 1 of the reaction takes place very much faster than stage 2, it can be assumed that AB and B react only when $a < b$, i.e. when the initial concentration of B is such that molecules of B surplus to the requirements of stage 1 are available.

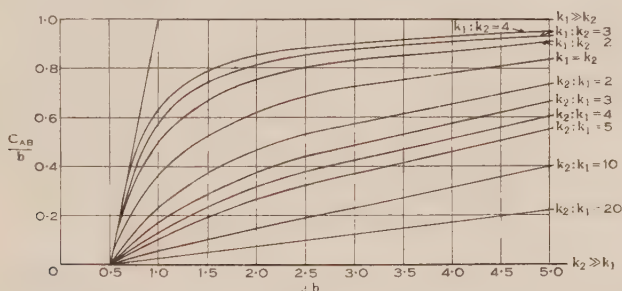


Fig. 2.—Dependence of final concentration of AB on initial concentration of A for various values of K (equations (13) and (12)).

Hence, when $a > b$,

$$C_{AB}=b; \quad C_{BAB}=0; \quad C_A=a-b, \quad \dots\dots\dots (16)$$

and when $b > a > b/2$,

$$C_{BAB}=b-a; \quad C_{AB}=2a-b; \quad C_A=0. \quad \dots\dots\dots (17)$$

Discussion and Conclusion

The theoretical final concentrations of the reactants and products have now been defined for all cases. In Figure 1 the final concentrations of the product AB and BAB are shown as functions of a , with b constant, for various special conditions pertaining to the kinetics of the two steps of the reaction. In Figures 2 and 3 the relationship between the final concentrations of the products and the initial concentrations of the reactants are shown for specific values of $k_2:k_1$. Experimental data can be conveniently interpreted against these curves. In applying the theory to the coupling of amino acids with diazonium compounds, the main interest will be centred on the final concentrations of the products, the azo-derivatives, which lend themselves to spectrophotometric investigation, but from the equations given, the final concentrations of the reactants can be expressed as required in a form similar to Figures 1-3.

The ratio of the velocity constants of the two consecutive steps, which can be determined by this method when $a \sim b/2$, can of course be interpreted in terms of the reactivity of the molecules A and AB . This has interesting implications concerning the effect of substitution in one position upon the energy of the molecule. In an organic molecule where the reaction with B leads to increased conjugation, this kinetic approach, used in conjunction with changes in the wavelength and intensity of absorption maxima, may provide a relatively simple method of contributing to the study of basic electronic phenomena within the molecule.

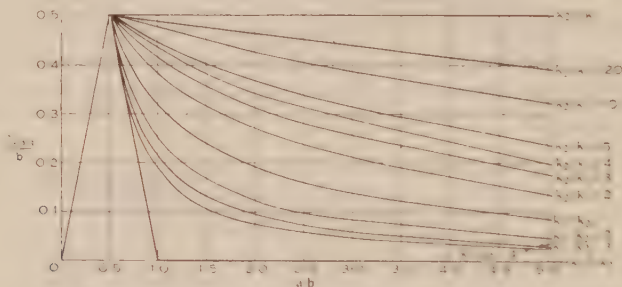


Fig. 3.—Dependence of final concentrations of BAB on initial concentration of A for various values of K (equations (14) and (12)).

Although this treatment is presented primarily to cope with problems arising in the study of the coupling of amino acids, free and in peptide linkage, with diazonium compounds, its kinetic implications should be applicable to the study of any reaction of the form specified in the Introduction, providing a state approximating to finality can be reached in an experimentally convenient period. Attention should be drawn, however, to a possible limitation in applying these results. The treatment assumes that the reactivity of the two positions on molecule A is *initially* the same. This will be so if the two points of reaction are symmetrically disposed, and may be true, or approximately so, in other cases. If, however, the initial reactivities differ appreciably, a more complex treatment is required.

THE CRYSTAL STRUCTURE OF POTASSIUM HEXACHLORORUTHENATE(IV)*

By C. S. ADAMS† and D. P. MELLOR†

Potassium hexachlororuthenate(IV), in which ruthenium has a magnetic moment of 3.07 Bohr magnetons, is the most strongly paramagnetic compound of the platinum metals yet described (Mellor 1943). This, together with the fact that there has been some confusion about the identity of the complex chloro-salts of ruthenium and the fact that the octahedral covalent radius of Ru^{IV} has not hitherto been determined, is the reason for briefly recording the results of an X-ray examination of the compound. The substance was prepared by the method described by Charonnat (1931).

As would be expected, the crystal belongs to the cubic system and is isomorphous with K_2PtCl_6 (Ewings and Pauling 1928). The length of the edge of the unit cell of K_2RuCl_6 , determined by the Bradley-Jay extrapolation method from powder photographs taken with $\text{CoK}\alpha$ radiation, is $9.738 \pm 0.001 \text{ \AA}$. The first 30 lines of the powder photograph are listed in Table 1. The parameter

TABLE 1

<i>hkl</i>	$\text{Sin}^2 \theta$		\sqrt{I}		<i>hkl</i>	$\text{Sin}^2 \theta$		\sqrt{I}	
	Obs.	Calc.	Obs.	Calc.		Obs.	Calc.	Obs.	Calc.
111	0.0247	0.0252	<i>s</i> ⁺	40.3	711	0.4293	0.4285	<i>m</i> ⁻	17.8
200	0.0338	0.0336	<i>m</i> ⁺	22.6	551				
220	0.0677	0.0672	<i>s</i>	27.7	640				
311	0.0932	0.0924	<i>m</i> ⁺	24.2	642	0.4710	0.4706	<i>w</i>	11.2
222	0.0997	0.1008	<i>s</i>	24.8	731	0.4966	0.4958	<i>w</i> ⁻	13.8
400	0.1353	0.1344	<i>s</i> ⁺	36.8	553				
331	0.1608	0.1597	<i>w</i>	15.6	800				
420	0.1694	0.1681	<i>w</i> ⁺	14.1	733	—	—	<i>absent</i>	1.3
422	0.2032	0.2017	<i>w</i> ⁺	15.6	820	0.5729	0.5714	<i>w</i>	10.7
511	0.2280	0.2269	<i>m</i>	20.4	644				
333					822				
440	0.2703	0.2689	<i>s</i>	28.7	660	0.6057	0.6050	<i>w</i>	12.2
531	0.2950	0.2941	<i>m</i> ⁻	16.6	751	0.6312	0.6302	<i>w</i>	13.2
600	0.3049	0.3025	<i>w</i>	13.6	555				
442					662				
620	0.3376	0.3361	<i>w</i>	11.3	840	0.6395	0.6386	<i>w</i> ⁻	10.9
533	0.3622	0.3613	<i>w</i> ⁻	1.7	911	0.6725	0.6722	<i>m</i>	19.2
622	0.3717	0.3697	<i>w</i>	13.4	753	0.6971	0.6974	<i>w</i>	14.0
444	0.4040	0.4033	<i>m</i> ⁻	16.6	842	0.7049	0.7058	<i>w</i> ⁻	5.6

* Manuscript received April 9, 1952.

† Chemistry School, University of Sydney.

a defining the Ru-Cl distance was determined by means of the method of inequalities and found to be 0.235 ± 0.005 , a value which leads to a reasonable agreement between observed and calculated intensities. The spots chosen for intensity comparisons were on rotation photographs taken with $\text{CuK}\alpha$ radiation. If one assumes a value of 0.99 \AA for the covalent radius of chlorine, the Ru-Cl distance of $2.29 \pm 0.04 \text{ \AA}$ leads to a value of $1.30 \pm 0.04 \text{ \AA}$ for the octahedral radius of Ru^{IV} . It is of interest to compare this value with the values obtained for covalent radii of the transition elements in different oxidation states, set out in Table 2 (Pauling 1939).

TABLE 2

Oxidation State	Fe	Co	Ni	Ru, Os	Rh, Ir	Pd, Pt	Ag, Au
II	1.23	1.32	1.39	1.33	1.43	1.50	
III		1.22	1.31	—	1.32	1.42	1.49
IV			(1.21)	1.30	1.29	1.31	1.41

The value shown in brackets is extrapolated. It will be noticed that as one goes diagonally across the table from left to right and down, as in the sequences Os^{II} , Ir^{III} , Pt^{IV} , and Ir^{II} , Pt^{III} , Au^{IV} , there is a regular decrease in radius. The uncertainty of the values of the radii quoted in Table 2, which is roughly the same throughout and about that of the uncertainty of the radius of Ru^{IV} , makes it debatable whether the regular decrease in the values for sequences of isoelectronic species is real.

The existence of such a decrease appears to be reasonable; if it is real, as it may be, Ru^{IV} fits into the sequence Co^{II} , Ni^{III} , Ru^{IV} .

On the assumption that the decrease is approximately 0.01 \AA in each step of the series, Ni^{II} appears to have an unusually large radius (1.39 \AA) in $[\text{Ni}(\text{NH}_3)_6]\text{Cl}_2$ (Wyckoff 1922) as compared with that of Rh^{IV} (1.29 \AA) (Dwyer, Nyholm, and Rogers 1947). Preliminary measurements indicate that the value for Ni^{II} is smaller and nearer to 1.33 \AA (Wunderlich, unpublished data). An attempt is also being made to determine the radius of Os^{III} in the compound $[\text{Os}(\text{NH}_3)_6]\text{ISO}_4$ which is isomorphous with the corresponding cobalt compound.

References

- CHARONNAT, R. (1931).—*Ann. Chim.* **16**: 5.
 DWYER, F. P., NYHOLM, R. S., and ROGERS, L. E. (1947).—*J. Roy. Soc. N.S.W.* **81**: 267.
 EWING, F. J., and PAULING, L. (1928).—*Z. Kristallogr.* **68**: 223.
 MELLOR, D. P. (1943).—*J. Roy. Soc. N.S.W.* **77**: 145.
 PAULING, L. (1939).—"Nature of the Chemical Bond," p. 169. (Cornell Univ. Press: Ithaca, New York.)
 WYCKOFF, R. W. (1922).—*J. Amer. Chem. Soc.* **44**: 1239.

ALKALOIDS OF THE AUSTRALIAN RUTACEAE: *GLYCOSMIS PENTAPHYLLA* (RETZ.) CORREA*

By A. W. MCKENZIE† and J. R. PRICE‡

Glycosmis pentaphylla (Retz.) Correa, a small tree or shrub assigned by Engler and Prantl (1931) to the tribe Aurantieae of the family Rutaceae and distributed from north-eastern Australia to Malaya and India, has been examined in the course of a survey of the occurrence of alkaloids in members of this family indigenous to Australia. The air-dried plant was found to contain approximately 0.01 per cent. of each of the two furoquinoline bases kokusaginine, 4,6,7-trimethoxyfuro-(2',3'-2,3)-quinoline (see Anet *et al.* 1952), and skimmianine, 4,7,8-trimethoxyfuro-(2',3'-2,3)-quinoline.

Experimental

Leaves and twigs (7.36 kg.), collected May 1951 at Chilli Creek, Cape York Peninsula, were milled and percolated with methanol (80 l.) at room temperature, the solvent evaporated, and the residue extracted 3 times with hot aqueous hydrochloric acid (5% ; 10 l. in all). The alkaloids were recovered from the acid solution, dissolved in methanol (250 ml.), and treated with excess methanolic picric acid, giving picrate *A*, 1.61 g. The base not precipitated was recovered from the methanolic picric acid mother liquors and, after another passage through 5% hydrochloric acid, dissolved in methanol (75 ml.) and again treated with excess methanolic picric acid, precipitating picrate *B*, 1.20 g. No more picrate was obtained by re-treating the mother liquors.

Picrate *A*, twice crystallized from ethanol, gave kokusaginine picrate, fine yellow needles, m.p. 217.5–218.5 °C.‡ alone or mixed with an authentic specimen.

Found: C, 49.2; H, 3.3; N, 11.3%.

Calculated for $C_{14}H_{13}O_4N.C_6H_3O_7N_3$: C, 49.2; H, 3.3; N, 11.5%.

Kokusaginine recovered from the picrate was converted to the characteristically sparingly soluble hydrochloride, m.p. 224 °C. (decomp.), and thence again to the base which separated from benzene as colourless prisms, m.p. 171–172 °C. undepressed by admixture with authentic kokusaginine.

Found: C, 65.0; H, 5.0; N, 5.5%.

Calculated for $C_{14}H_{13}O_4N$: C, 64.9; H, 5.0; N, 5.4%.

The R_F values for butanol-5% aqueous acetic acid were as follows:

Authentic kokusaginine 0.80

Kokusaginine from *G. pentaphylla* 0.80.

Additional confirmation was furnished by heating with methyl iodide in a sealed tube at 100 °C. for 1 hour giving isokokusaginine, m.p. and mixed m.p. with an authentic specimen 250–251 °C.

Picrate *B* crystallized from methanol as yellow needles, m.p. 197–198 °C. undepressed by admixture with skimmianine picrate.

Found: C, 49.2; H, 3.4; N, 11.7%.

Calculated for $C_{14}H_{13}O_4N.C_6H_3O_7N_3$: C, 49.2; H, 3.3; N, 11.5%.

* Manuscript received March 21, 1952.

† Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

‡ Melting points are corrected; microanalyses are by the C.S.I.R.O. Microanalytical Laboratory.

The base recovered from the picrate crystallized from aqueous ethanol as colourless prisms, m.p. 177–178 °C. alone or mixed with an authentic specimen of skimmianine.

Found: C, 65.1; H, 5.0; N, 5.7%.

Calculated for $C_{14}H_{15}O_4N$: C, 64.9; H, 5.0; N, 5.4%.

The R_F values for butanol-5% aqueous acetic acid were as follows:

Authentic skimmianine	0.88
Skimmianine from <i>G. pentaphylla</i>	0.87.

The authors are indebted to Professor E. J. Hartung for providing accommodation and facilities in the Chemistry Department, University of Melbourne, to Mr. G. K. Hughes for specimens of kokusaginine and some derivatives, and to Mr. L. J. Webb, Division of Plant Industry, C.S.I.R.O., for supplying the material used in this investigation.

References

- ANET, F. A. L., GILHAM, P. T., GOW, P., HUGHES, G. K., and RITCHIE, E. (1952).—*Aust. J. Sci. Res. A* 5: 412.
- ENGLER, A., and PRANTL, K. (1931).—“Die Natürlichen Pflanzenfamilien.” Vol. 19a. (Wilhelm Engelmann: Leipzig.)

THE ABSOLUTE MEASUREMENT OF HIGH-ENERGY RADIATION

By B. M. SPICER*

[Manuscript received June 19, 1952]

Summary

The relationship between incident intensity of X-rays and the ionization produced by them in a small cavity behind an aluminium wall of given thickness has been derived, and calculations performed for a specific case. The conditions of validity of Gray's relation are the only ones imposed. The calculation was performed for monoenergetic radiation, and extended to the case of a bremsstrahlung spectrum of the form given by Schiff (Adams 1948). Comparison is made with the work of Fowler, Lauritsen, and Lauritsen (1948) and Lawson (1950).

I. INTRODUCTION

The application of the bremsstrahlung from a betatron or synchrotron to the study of photonuclear reactions has emphasized the need for improvement in the absolute measurement of the flux of high-energy radiation, because knowledge of the number of photons striking the target material is fundamental in the determination of the cross section for the reaction.

The most satisfactory method of measuring high-energy X-rays is to use the cavity ionization chamber (International Commission on Radiological Units 1951). The relation between energy absorbed per unit volume in the solid wall and the ionization produced per unit volume of a small cavity was treated first by Gray (1936). This relation was derived under the conditions that (a) the wall thickness was greater than the range of secondary electrons produced by the X-rays, (b) the atomic number of the wall was nearly equal to that of the cavity gas, and (c) the secondary electrons lost a negligible amount of energy in crossing the cavity. Under these conditions, Gray found that

$$J_A = \frac{E_s}{w\varphi}, \quad \dots\dots\dots (1)$$

where J_A is the number of ion pairs produced per unit volume of the cavity,

E_s is the energy absorbed per unit volume of the solid,

w is the energy required to produce one ion pair in the cavity gas, and

φ is the relative stopping power of the wall compared with the gas.

In the present paper the variation of E_s with incident quantum energy has been calculated for an X-ray flux of 1 quantum/cm.², and the relation between ionization within the cavity and the incident X-ray flux is then obtained from the equation (1). These results are applied to the case of an aluminium-walled ionization chamber used for the absolute measurement of 15 MeV. X-radiation.

* Physics Department, University of Melbourne.

Results of similar calculations have been published by Lawson (1950) and Johns *et al.* (1950), but in both cases no mention was made of the method used to calculate the energy absorbed per unit volume of wall. Fowler, Lauritsen, and Lauritsen (1948) discussed the same problem, considering only the ionization due to the Compton electrons and pairs produced in the walls. These considerations have been extended to include the effects due to (a) secondary Compton effect, which will increase the ionization by up to 10 per cent., (b) absorption of X-rays in the chamber walls, and (c) annihilation of positrons while in motion. The effects of (c) decrease the ionization due to positrons by about 10 per cent. at 15 MeV. Its importance increases slowly with X-ray energy above 5 MeV. The photoelectric effect is ignored for X-ray energies greater than 1 MeV., and material having atomic number less than 14. Radiation loss by secondary electrons makes no appreciable contribution for X-ray energies up to 15 MeV., and is therefore neglected in this range.

II. ENERGY ABSORBED PER UNIT VOLUME OF ION CHAMBER WALL - MONOENERGETIC X-RAYS

(a) *Energy Absorbed per Unit Volume from Secondary Electrons*

Consider a parallel beam of monoenergetic quanta, energy μ MeV., $W = \gamma mc^2$, incident on a block of material of atomic number Z . Let there be in this block a cavity at depth D , where D is at least equal to the maximum range of secondary electrons produced by the quanta.

Suppose the incident beam to consist of 1 quantum cm^{-2} . Let the number of negative electrons formed by Compton effect and pair production in a thin lamina of width dx at depth x in the block, and having kinetic energy between T and $T + dT$, be

$$n_e(T, W) \cdot dT \cdot dx \cdot e^{-\mu x}, \dots\dots\dots (2)$$

where T is the kinetic energy, and μ is the linear absorption coefficient in the wall material for X-rays of W MeV. energy. The factor $e^{-\mu x}$ allows for the absorption of quanta in the chamber wall.

The negative electrons formed in the small element will proceed almost entirely in the forward direction, and will lose energy uniformly along their path, that is, $-dT/dy = T/R = B$. This is justified by the linearity of the curve of average range, R , *v.* energy (Fowler, Lauritsen, and Lauritsen 1948). The energy lost in a small distance dy at depth y ($y > x$) by the electrons formed in dx is given by

$$-\frac{dT'}{dy} \cdot dy \cdot n_e(T, W) \cdot e^{-\mu x} \cdot dT \cdot dx, \dots\dots\dots (3)$$

where the energy loss in distance $y - x$ is $T - T' = T(y - x)/R$. The energy lost in an interval dy by all the electrons of kinetic energy between T and $T - dT$ which reach dy is given by

$$\begin{aligned} dE_e &= dy \cdot n_e(T, W) dT \cdot \int_{y-R}^y e^{-\mu x} B \cdot dx \\ &= dy \cdot n_e(T, W) \cdot dT \cdot \frac{e^{-\mu y}}{\mu} \cdot (e^{\mu R} - 1) \cdot B. \dots\dots\dots (4) \end{aligned}$$

The total energy lost in dy by all electrons which reach depth y is given by

$$E_e = dy \cdot \frac{e^{-\mu y}}{\mu} \cdot \int_0^{T_m} n_e(T, W) \cdot B \cdot (e^{\mu R} - 1) dT. \quad \dots\dots\dots (5)$$

The value taken for T_m is the value for the maximum energy given to an electron in the Compton effect ($T_m = 2\gamma W / (1 + 2\gamma)$). Because of the complicated dependence of $n_e(T, W)$ on T , the integration must be performed numerically.

Values of $n_e(T, W)$ were obtained by summing the formulae for Compton effect and pair production given by Rossi and Griesen (1941). The probability

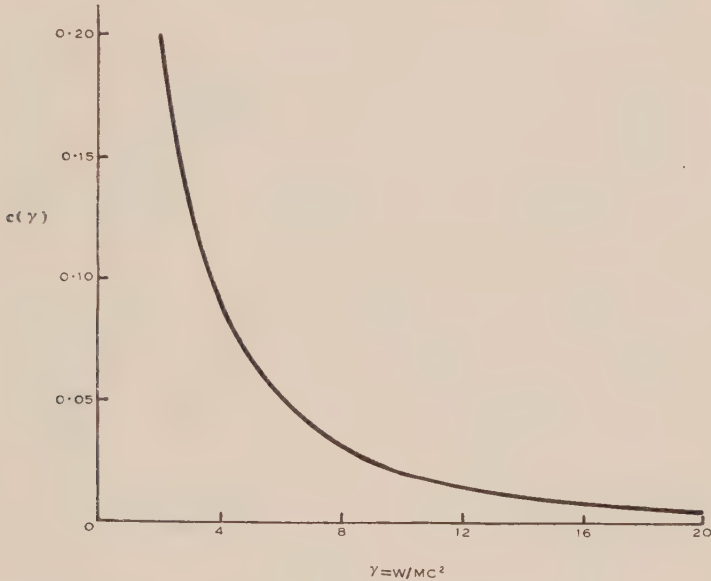


Fig. 1.—The function $c(\gamma)$.

for a photon of energy W , traversing thickness dx , to undergo a Compton collision in which the scattered photon has energy between W' and $W' + dW'$ is given by

$$K(W, W') dW' \cdot dx = \pi N Z r_0^2 \frac{dW'}{\gamma \cdot W'} \left[1 + \left(\frac{W'}{W} \right)^2 - \frac{W'}{W} \sin^2 \theta \right] dx,$$

where $W' = W - T$, N is the number of atoms/cm.³, and

$$\sin^2 \theta = 2(W - W') / \gamma W' + (W - W')^2 / \gamma^2 W'^2.$$

The probability of a photon of energy W , traversing thickness dx , producing an electron-positron pair in which the electron has fractional energy between $u = (T + mc^2) / W$ and $u + du$, is given by

$$\Psi(W, u) du \cdot dx = 4\alpha N Z r_0^2 du [u^2 + (1 - u)^2 + \frac{2}{3}u(1 - u)] [\ln\{2\gamma u(1 - u)\} - \frac{1}{2} - c(\gamma)] dx,$$

where $\alpha = 2\pi e^2 / \hbar c = 1/137$ and $c(\gamma)$ is the function shown in Figure 1.

(b) Energy Absorbed per Unit Volume from Secondary Positrons

Positrons formed in the pair production process will proceed predominantly in the forward direction and will lose energy in the same manner as the electrons.

Further, the positrons may be annihilated while in motion, and the intensity of the beam of secondary positrons in the wall material is thereby lowered.

By analogy with the case of negative electrons, the energy lost in dy by positrons of kinetic energy between T and $T+dT$ is

$$dE_p = B \cdot dy \cdot n_p(T, W) \cdot dT \cdot \int_{y-R}^y e^{-\mu x} (1-Q) dx, \quad \dots\dots\dots (6)$$

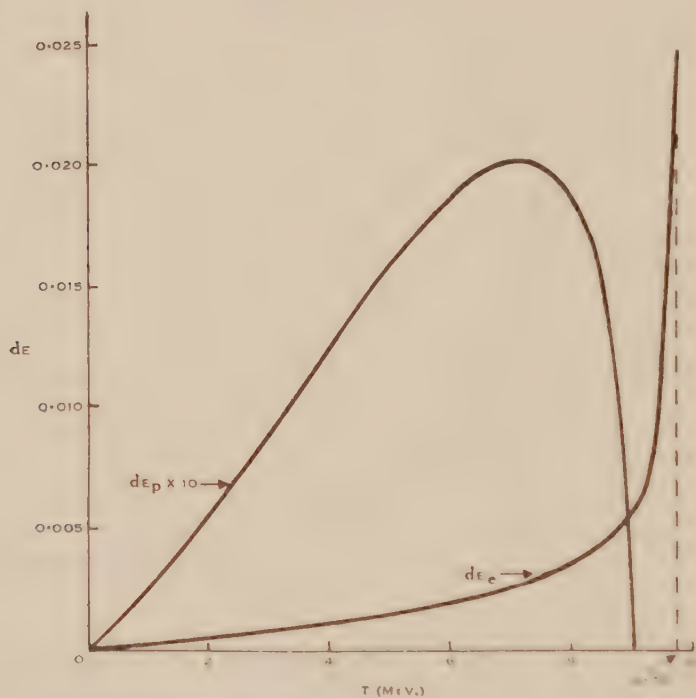


Fig. 2.— dE_p and dE_e plotted against T for $W=10$ MeV.

where Q is the probability that a positron of initial total energy E will be annihilated in distance $y-x$,

$$Q = \frac{1}{T - T_1} \int_{T_1}^T w(T') \cdot dT',$$

where T is the kinetic energy of positron at dy , and $w(T')$ is the probability of annihilation of a positron per unit energy interval (Heitler 1936, p. 230). Figure 2 shows curves of dE_e and dE_p plotted against T .

We approximate to $w(T')$ by the function $0.1168 - 0.0079T'$. This is a crude approximation, but it makes less than 1 per cent. error in the final value of E_p . We find $Q = 0.1168 - 0.0039 \cdot B(2R - y + x)$.

Then

$$dE_p = B \cdot dy \cdot \frac{e^{-\mu y}}{\mu} \cdot dT \cdot n_p(T, W) \mu R (a + bR),$$

where $a=0.8832+0.0079T$, $b=-0.004B$, and $e^{\mu R}$ is replaced by $1+\mu R$. This introduces an error of less than 1 per cent.

Integration over T gives

$$E_p = \frac{e^{-\mu y}}{\mu} \cdot dy \cdot \int_0^{T_p} n_p(T, W) \cdot \mu T \cdot [0.8832 + 0.0039T] \cdot dT, \quad \dots\dots\dots (7)$$

where T_p is the maximum kinetic energy given to a positron in the pair production process.

(c) *Energy Absorbed per Unit Volume Due to a Second Compton Effect*

This analysis must be performed by averaging over all secondary quanta. For an incident X-ray intensity of 1 quantum/cm.², the number of quanta scattered from the primary beam by Compton effect in a thickness dx at depth x in the wall material, is $\sigma_1 dx N e^{-\mu_1 x}$, where σ_1 =total cross section for Compton scattering from the primary X-ray beam by one electron. The subscripts "1" refer to the primary photon beam, subscripts "2" to the secondary photon beam, etc.

If W is the energy of the primary photons, the amount of energy lost from the X-ray beam is $W\sigma_{1a}$, where σ_a is the cross section for energy absorption by Compton scattering. If $\gamma=W/(mc^2)$, we have (Cave, Corner, and Liston 1950)

$$\sigma = \frac{\pi r_0^2}{\gamma^2} \left[\left(\gamma - 2 - \frac{2}{\gamma} \right) \ln(1+2\gamma) + 4 + \frac{\gamma}{2} - \frac{\gamma}{2}(1+2\gamma)^2 \right], \quad \dots\dots\dots (8)$$

$$\sigma_a = \pi r_0^2 \left[(\gamma^{-1} - 2\gamma^{-2} - 3\gamma^{-3}) \ln(1+2\gamma) - \frac{8\gamma^2}{3(1+2\gamma)^3} + \frac{6+22\gamma+18\gamma^2-2\gamma^3}{\gamma^2(1+2\gamma)^2} \right] \quad \dots\dots\dots (9)$$

The average energy carried away per secondary photon is

$$W' = \frac{\sigma_1 - \sigma_{1a}}{\sigma_1} \cdot W.$$

These photons have a certain probability of being scattered again and producing secondary electrons which will lose energy in the wall and so contribute to the energy absorbed per unit volume of solid. We assume that these quanta of energy W' are absorbed only by a Compton effect. Pair production is neglected for these quanta because, as W ranges from 2 to 15 MeV., W' varies from 0.9 to 4.5 MeV. Photoelectric effect may still be neglected.

The number of secondary quanta scattered in thickness dx will have a probability σ_2 per electron of wall material of being scattered again. The number of secondary quanta further scattered in a distance dy at total depth y is given by $(\sigma_2 dy N)$ times the flux of secondary quanta. The electrons produced by these quanta will have average kinetic energy $W' \cdot \sigma_{2a}/\sigma_2$, where W' is the average energy of secondary quanta.

Energy loss per unit path length of these electrons will be B , assuming again the proportionality of range and energy, that is, $B = (\sigma_{2a}/\sigma_2) \cdot (W'/R)$.

The total energy absorbed in thickness dz at depth D from electrons produced in secondary Compton effect is

$$\frac{\sigma_{2a}}{\sigma_2} \cdot \frac{W'}{R} \cdot dz \cdot \sigma_2 \cdot dy \cdot N^2 \cdot \sigma_1 dx \cdot e^{-\mu_1 x} \cdot e^{-\mu_2(y-x)}.$$

The total energy lost by ionization in dz for the region $x=0$ to D is obtained by integrating this expression between the limits $x=0, D$ and $y=D-R, D$

$$E_c = B \cdot dz \cdot \int_0^D \int_{D-R}^D N^2 \sigma_1 \sigma_2 e^{-\mu_1 x} e^{-\mu_2(y-x)} \cdot dy dx.$$

We then have

$$E_c = \frac{\sigma_{2a}(\sigma_1 - \sigma_{1a})}{(\mu_2 - \mu_1)} \cdot \frac{W}{R} \cdot dz \cdot N^2 \cdot (e^{-(\mu_1 + \mu_2)D} - 1) e^{-\mu_1 D}, \quad \dots (10)$$

if we replace $e^{\mu_2 R}$ by $1 + \mu_2 R$.

(d) Total Energy Absorbed per Unit Volume

The total energy absorbed per unit volume of wall material at depth D is given by

$$E_s = E_e + E_p + E_c = \frac{e^{-\mu D}}{\mu} [f(W) + g(W)] + E_c,$$

where $f(W)$ is given in equation (5), $g(W)$ in equation (7), and E_c by equation (10). The functions f and g are dependent only on the properties of the secondary electrons, and are therefore independent of the properties of the ion chamber wall.

(e) Numerical Results

Since the object of this calculation was to give results for 15 MeV. bremsstrahlung, the values of E_s have been calculated for X-ray energies up to that value. The ionization chamber is assumed to have a 2.54 cm. thick aluminium wall. Results are shown in Tables 1 and 2. The calculation of E is separated from that of E_e and E_p .

TABLE 1
TOTAL ENERGY ABSORBED PER PHOTON/CM.² IN EACH CM.³ OF WALL AT A DEPTH OF 2.54 CM.

W (MeV.)	T_p (MeV.)	μ (cm. ⁻¹)	$f(W)$	$g(W)$	$e^{-\mu D}/\mu^*$ (cm.)	$E_e + E_p$ (MeV./cm. ²)
2	1.773	0.115	0.00695	—	6.426	0.0447
4	3.760	0.086	0.0162	0.00079	9.275	0.1577
6	5.755	0.071	0.0170	0.00206	11.684	0.2232
8	7.752	0.065	0.0207	0.00368	12.970	0.3165
10	9.751	0.061	0.0222	0.00529	13.965	0.3840
12	11.750	0.061	0.0256	0.00741	13.965	0.4609
14	13.749	0.061	0.0298	0.00975	13.965	0.5330
15	14.749	0.060	0.0296	0.0110	14.098	0.5724

* The fact that not all secondaries are emitted at 0° with the incident X-ray direction will result in a slight increase in the effective wall thickness. This has been corrected for on an average by taking the wall thickness as $D \cos \tau$, where τ is the average angle of emission of the electrons.

TABLE 2
 THE SECONDARY COMPTON EFFECT

W (MeV.)	σ_1 (10^{-26}cm.^2)	σ_{1a} (10^{-26}cm.^2)	W' (MeV.)	σ_{2a} (10^{-26}cm.^2)	E_c (MeV./cm. ²)
2	14.692	7.810	0.937	9.242	0.00937
4	9.623	5.837	1.574	8.273	0.0108
6	7.499	4.729	2.217	7.409	0.0117
8	6.006	4.005	2.665	6.939	0.0109
10	5.112	3.476	3.201	5.949	0.0099
12	4.469	3.109	3.651	5.764	0.0102
14	3.982	2.810	4.121	5.670	0.0099
15	3.806	2.683	4.426	5.475	0.0100

III. IONIZATION PRODUCED PER UNIT VOLUME IN THE CAVITY

Gray's relation is now used to relate the total energy absorbed per unit volume in aluminium to the ionization in the air of the cavity. The values of ρ to be used were estimated from the Bethe-Bloch formula (Heitler 1936, p. 219) and the average energy of secondary electrons. The Bethe-Bloch formula gives the correct value for ρ , even though the Halpern-Hall correction (Halpern and Hall 1948) has not been applied. This is so because the ratio of the corrections for aluminium and air approximates to the ratio of energy loss calculated from the Bethe-Bloch formula to within 1 per cent. The average electron energy \bar{E} , and the values of ρ are shown in Table 3, which also shows the results of calcula-

 TABLE 3
 THE NUMBER OF ION PAIRS PER CM.³ OF CAVITY PER QUANTUM/CM.² INCIDENT ON
 A 2.54 CM. ALUMINIUM WALL

W (MeV.)	$\bar{E} = T + mc^2$ (MeV.)	ρ	E_s (MeV./cm. ³)	J_A (ion pairs/cm. ³)
2	1.55	1897	0.0541	0.90
4	3.42	1911	0.1685	2.71
6	5.1	1917	0.2349	3.77
8	6.8	1919	0.3274	5.25
10	8.5	1922	0.3939	6.31
12	10.2	1924	0.4711	7.54
14	11.9	1926	0.5629	9.00
15	12.75	1926	0.5824	9.31

tion and the number of ion pairs formed per cm.³ behind a 2.54 cm. aluminium wall for an X-ray intensity of 1 quantum/cm.² incident on the outer wall of the chamber. The value 32.5 eV. is used for w (Gray 1936).

IV. ACCURACY OF CALCULATION

The results given for the value of E_s will be accurate only for X-ray energies higher than 4 MeV. The lower limit is set by the fact that the linear relation

between range and energy of electrons breaks down below about 1.5 MeV. Below 1 MeV., the rate of energy loss of electrons increases with fall of energy. Since for X-rays of energy below 5 MeV., the proportion of secondary electrons having energy below 1.5 MeV. is not insignificant, the results for 2 and 4 MeV. are expected to be too low.

Radiation loss by the electrons has been neglected. Since we are dealing with X-ray energies up to 15 MeV., where the energy loss of electrons by radiation does not exceed 5 per cent. of the total energy loss, it may be safely neglected. Its inclusion in E_c introduces a discrepancy of less than 1 per cent.

Taking all factors into consideration the calculations are considered accurate to within 5 per cent., except for the cases of 2 and 4 MeV. quantum energies. By far the greatest errors are introduced in the assumption that R is proportional to T , and in the evaluation of the integrals. The error in the latter is no more than 2 per cent.

The calculation may be taken to higher energies in a way which involves calculation of $f(W)$ and $g(W)$ for the higher values of W , and a scaling of the factor $e^{-\mu D}$ using the fact that D must be greater than the average range of the most energetic secondary electrons.

V. EXTENSION FOR A CONTINUOUS ENERGY SPECTRUM OF X-RAYS

Because the X-rays in a continuous energy spectrum will not produce the same number of ion pairs per incident quantum, the interpretation of data will necessarily hinge on the average value for the number of ion pairs formed per incident quantum. This quantity is determined from the formula

$$\bar{J}(W_m) = \frac{\int_0^{W_m} J(W) \cdot p(W, W_m) \cdot dW}{\int_0^{W_m} p(W, W_m) \cdot dW}, \quad \dots\dots\dots (11)$$

where $J(W_m)$ is the average number of ion pairs formed per cm.² of cavity per incident quantum for a continuous spectrum of maximum energy W_m ,

$J(W)$ is the number of ion pairs formed per incident quantum of energy W (the value of this quantity is taken from Table 4), and

$p(W, W_m)$ is the probability that a quantum, in a continuous spectrum of maximum energy W_m , will have energy between W and $W + dW$.

For betatrons and synchrotrons in which the X-ray beam is produced by the electrons striking a flag target not more than 0.010 in. thick, the energy spectrum follows closely the theoretical form given by Schiff (Adams 1948). The intensity in the forward direction is proportional to dp/dW , where

$$\frac{dp}{dW} = \frac{1}{W} [2(1-z)(\ln \alpha - 1) + z^2(\ln \alpha - \frac{1}{2})], \quad \dots\dots\dots (12)$$

and

$$1/\alpha^2 = 1/\alpha_1^2 + 1/\alpha_2^2,$$

$$\alpha_1 = \frac{2E_0(1-z)}{mc^2z},$$

$$\alpha_2 = 191/Z^{1/3},$$

$$z = W/W_m,$$

W_m = total energy of electron,

Z = atomic number of target material.

$\bar{J}(W_m)$ has been calculated for the Schiff intensity spectrum with W_m in the range 10–15 MeV., and the results were found to fit the formula $\bar{J}(W_m) = 0.158 W_m - 0.05$ in the region stated. From this mean value and the total ionization measured, the number of quanta incident on the chamber can be calculated. The number per unit energy interval may then be calculated from equation (12).

VI. DISCUSSION

The results given in the last column of Table 3 lie nearly on the straight line given by

$$J_A = 0.625W + 0.140,$$

which is obtained from a least squares fit of the calculated points. The values given by this equation are shown in Table 4, along with the results of Fowler,

TABLE 4
COMPARISON OF RESULTS WITH THOSE OF OTHER AUTHORS

W (MeV.)	J_A (Author)	J_A (Lawson)*	J_A (Fowler, Lauritsen, and Lauritsen)†
2	1.39	1.62	1.40
4	2.64	2.75	2.59
6	3.89	4.20	3.82
8	5.14	5.50	5.06
10	6.39	6.62	6.22
12	7.64	7.64	7.62
14	8.89	8.80	8.81
15	9.51	9.43	9.13

* The results attributed to Lawson have been obtained by extrapolating results given for wall thicknesses of 5, 7.5, and 10 cm. Their accuracy may therefore be questioned, but it should be much better than 10 per cent.

† The results of Fowler, Lauritsen, and Lauritsen are multiplied by $e^{-\mu D}$ to allow for the absorption of X-rays in the chamber walls.

Lauritsen, and Lauritsen (1948) and Lawson (1950). In both these cases there have been corrections applied so that the results apply to the same conditions of ionization.

Although the other authors quoted in Table 4 give results which are close to those derived here, there are several differences both in method of calculation and numerical values taken for the various quantities.

The energy absorbed per unit volume of the solid was calculated by Fowler, Lauritsen, and Lauritsen (1948) without taking into account the secondary Compton effect. Thus their values are, in the main, lower than the others. The accuracy of their results improves as the X-ray energy is raised, owing to the decreasing importance of the secondary Compton effect.

The results of Johns *et al.* (1950) are not easily compared with those given here as they used (a) lucite as the medium for production of secondary electrons, and (b) a totally different method of allowing for ionization due to secondary effects. The stopping power for electrons of lucite relative to air is given, but no allowance is made for its variation with energy. No data on the method of calculation are given.

Laughlin (1952) has measured the energy flux of radiation from a betatron operating at 22.5 MeV. by a calorimetric method, and has related this to ionization measurements in the same beam. This result is then used to calculate the number of photons per cm.²-roentgen per unit interval of the energy spectrum. Results obtained using the present calculation will decrease with increasing energy more rapidly than do those of Laughlin. This same trend is seen in results quoted by Laughlin from Johns *et al.* (1950).

The effect due to annihilation of positrons while in motion is to decrease E by about 10 per cent. The energy loss of secondary electrons by bremsstrahlung is negligible, but would not be so at 20 MeV.

The value of φ , the relative stopping power of wall to gas, is taken here from the Bethe-Bloch formula. The φ calculated therefrom is in agreement with experiment (Hereford 1948). Lawson uses the approximate formula given by Rossi and Griesen (1941), but Fowler, Lauritsen, and Lauritsen (1948) develop their own approximation for φ . Both of these are in error by several per cent. at 10 MeV. Further, Lawson uses a value for the energy to produce an ion pair of 32 eV. Gray (1936) and recent reviewers (Gray 1944; Wang 1950) give 32.5 eV.

In view of these differences, the agreement between the three sets of results is to some extent fortuitous. In the case of Lawson's work, the agreement comes almost entirely from the cancellation of differences in the values of ε and w .

VII. ACKNOWLEDGMENTS

The author wishes to thank Professor L. H. Martin for continued interest in the work; Mr. E. Muirhead for many helpful discussions; and Dr. W. B. Lasich for suggesting the problem.

VIII. REFERENCES

- ADAMS, G. D. (1948).—*Phys. Rev.* **74**: 1707.
 CAVE, L., CORNER, J., and LISTON, R. H. (1950).—*Proc. Roy. Soc. A* **204**: 223.
 FOWLER, W. A., LAURITSEN, C. C., and LAURITSEN, T. (1948).—*Rev. Mod. Phys.* **20**: 236.
 GRAY, L. H. (1936).—*Proc. Roy. Soc. A* **156**: 578.

- GRAY, L. H. (1944).—*Proc. Camb. Phil. Soc.* **40**: 72.
- HALPERN, O., and HALL, H. (1948).—*Phys. Rev.* **73**: 447.
- HEITLER, W. (1936).—"The Quantum Theory of Radiation." (Clarendon Press: Oxford.)
- HEREFORD, F. L. (1948).—*Phys. Rev.* **74**: 574.
- INTERNATIONAL COMMISSION ON RADIOLOGICAL UNITS (1951).—*Nucleonics* **8**: 28.
- JOHNS, H. E., KATZ, L., DOUGLAS, R. A., and HASLAM, R. N. (1950).—*Phys. Rev.* **80**: 1062.
- LAUGHLIN, J. S. (1952).—*Brit. J. Radiol.* **25**: 12.
- LAWSON, J. D. (1950).—Atomic Energy Research Establishment Rep. G/R 555.
- ROSSI, B., and GRIESEN, K. (1941).—*Rev. Mod. Phys.* **13**: 240.
- WANG, T. W. (1950).—*Nucleonics* **7**: 55.

THE ABSORPTION AND REFLECTION OF MICROWAVE RADIATION BY A MERCURY-VAPOUR DISCHARGE

By H. A. PRIME*

[*Manuscript received March 3, 1952*]

Summary

The power transmission and reflection coefficients of a limited volume of ionized gas (a mercury-vapour discharge) located within a waveguide have been measured at a frequency of approximately 10,000 Mc/s. by a microwave method. From these coefficients the real and imaginary components of the complex conductivity of the discharge are evaluated. The results show that the real component of the conductivity σ_r is a linearly increasing function of the discharge current, whereas the rate of increase of the imaginary component σ_i , which is negative, decreases with increasing discharge current. The ratio σ_i/σ_r decreases with increase in current, but is of the order of unity due to the fact that the gas pressure is sufficiently high (~ 1 atm.) to make σ_r comparable with σ_i . The theoretical basis of the work is presented in an appendix in which the particular case of high pressure conditions is discussed.

I. INTRODUCTION

The work described in this paper concerns an experimental study of the absorption and reflection of electromagnetic radiation (wavelength ~ 3 cm.) by a D.C. sustained discharge in mercury vapour. In addition, certain theoretical considerations relating to the work are discussed in Appendix I.

The basis of the method described is the measurement of the power transmission and reflection coefficients of a limited volume (or "slab") of ionized gas located within a rectangular waveguide operating in the H -dominant mode. From the theoretical point of view, the fundamental theme is the solution of a given boundary-value problem in the propagation of electromagnetic waves through a dispersive medium. In particular, the experimental results have been analysed to determine the variation of the complex conductivity of the ionized gas with changes in the current through the discharge, and the microwave properties of the discharge are discussed in the light of this analysis.

One further point must be emphasized, namely, that the ionization of the gas is controlled from a D.C. source. The microwave radiation plays no part in sustaining the discharge, because the electromagnetic field is insufficient to raise the electron energies to inelastic collision levels. The radiation is used only for investigating the characteristics of the discharge in the same way that a probe might be inserted to investigate the distribution of electron energies.

* Electrical Engineering Department, University of Adelaide.

II. EXPERIMENTAL APPARATUS AND METHOD

A block diagram of the microwave test assembly is shown in Figure 1.

The section of guide containing the discharge tube is shown in Figure 2 (a). As the diameter of the discharge tube was greater than the width of the guide across the broad face, shunt or *H*-plane T-junctions were fitted on each side of the guide and opposite to each other, and a hole drilled through the resulting "crossover" system to take the discharge tube as shown in the figure. Each shunt element was fitted with a short-circuiting piston, and the positions of these pistons could be adjusted.

The discharge tube employed for the measurements was obtained from a street-lighting unit, the outer safety jacket first being removed to leave only the tube itself and connecting leads, as shown in Figure 2 (b). During the

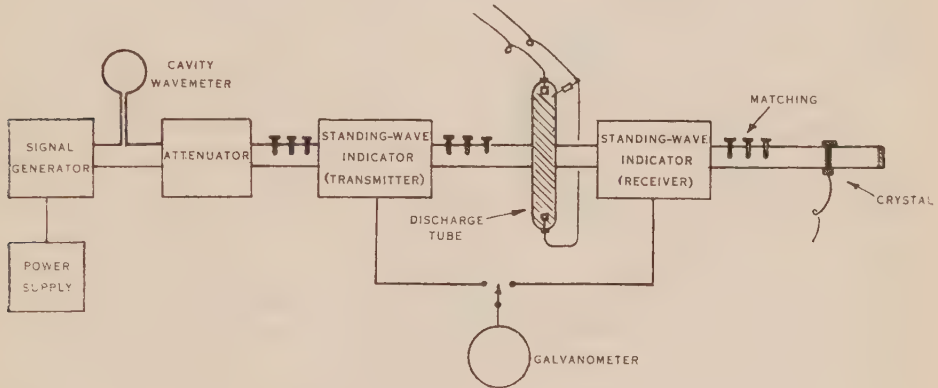


Fig. 1.—Microwave test assembly.

warming-up period, the tube is at first filled with a blue haze (due to traces of argon), but, as the temperature and pressure rise, the discharge contracts to a chord of light having an intense blue-green colour; as the current is increased this chord gradually expands in diameter until it fills the entire tube cross section. The final operating pressure is usually between 1 and 5 atm. The discharge tube was clamped in position in the crossover section, so that the centre of the discharge region was irradiated by the microwave power.

A number of preliminary adjustments of the apparatus were necessary before readings could be taken. The attenuator setting was adjusted until variation of matching components or the discharge-tube current did not affect the output power of the klystron and it could be concluded that the klystron was working effectively into a matched load. The positions of the two pistons at the crossover section were adjusted to give maximum current at the terminating crystal. It is difficult to assess the exact conditions existing at the discharge tube when the above adjustment is made, owing to the large number of evanescent modes which must undoubtedly obtain at the crossover region. However, as the crystal current did not appear to be unduly sensitive to movement of the pistons, their adjustment was not critical and once made could be left; nevertheless, it was convenient to assume that for maximum crystal current there

were virtual short circuits across the planes *ab* and *cd* of Figure 2 (*a*). Owing to the large diameter of the tube, there was inevitably leakage of power from the crossover region in the direction of the tube axis; however, such power was rapidly attenuated by the discharge and hence, for the measurements, it could be lumped with the power absorbed by the discharge within the waveguide cross section to give a resultant degree of absorption. The match between the crystal and guide assembly was maintained by adjustment of the final matching section to keep the standing-wave ratio on the receiver side less than 1.15, so that the crystal current was proportional to the power transmitted through the discharge tube (assuming square-law crystal operation).

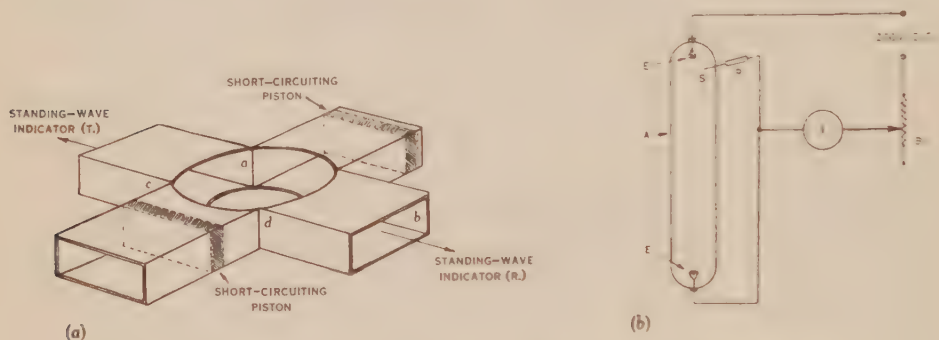


Fig. 2 (*a*).—Crossover assembly. Fig. 2 (*b*).—Discharge tube.

A, glass envelope; *B*, ballast resistor; *E*, electrodes; *I*, ammeter; *R*, starting resistor; *S*, starting electrode.

The experiments consisted of the measurement of the transmitted power P_t (proportional to the galvanometer deflection) and the value of the standing-wave ratio on the transmitter side s as a function of the discharge-tube current I_p for various initial experimental conditions. In Experiment 1 the system was matched by the section immediately preceding the discharge tube before the discharge tube was inserted (initial value of $s = 1.02$). For Experiment 2 the system was matched after the insertion of the tube (initial $s = 1.15$); whereas for Experiment 3 the system was rematched at every stage of the experiment. Thus, corresponding to each value of I_p , two values of P_t and s were noted: that resulting from the preceding change of I_p and that obtaining after the rematching adjustment. The results show that the three experiments are complementary from the point of view of attenuation calculations.

III. EXPERIMENTAL RESULTS

(*a*) Determination of T_p and η

In considering the results it is necessary to introduce certain parameters. Accordingly, the following ratios are defined:

$$\eta = P_r/P_i, \quad T_p = P_t/P_i, \quad \dots \dots \dots (1)$$

where P_i is the power incident on the discharge-tube section, P_r is the power reflected by the discharge tube alone or by the combination of tube and discharge

when the latter is present, and P_t is the power transmitted through the discharge and ultimately detected by the crystal. It can also readily be established that

$$\eta = \left(\frac{s-1}{s+1} \right)^2, \quad \dots\dots\dots (2)$$

where s is the voltage standing-wave ratio in the incident medium. Account must also be taken of the power absorbed either by the glass walls of the discharge

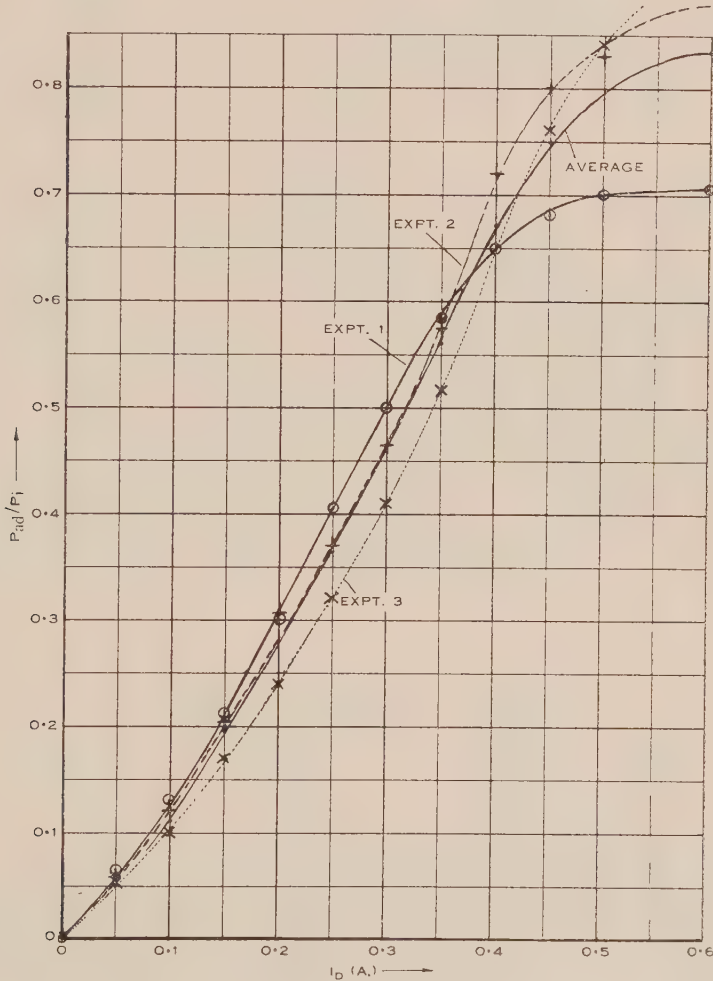


Fig. 3.— P_{ad} against I_D .

tube P_{ag} or by the discharge itself P_{ad} (included in P_{ad} is that power absorbed by the discharge external to the waveguide due to escape through the crossover hole). Assuming the guide to be loss-free, then

$$P_i = P_r + P_t + P_{ag} + P_{ad}$$

or

$$P_{ad}/P_i = 1 - \eta - T_p - P_{ag}/P_i. \quad \dots\dots\dots (3)$$

As the waveguide system is matched to both the source and the termination, the parameter T_p represents the insertion loss of the obstacle and, as indicated by equation (3), is made up of reflection loss by virtue of the standing wave on the incident side of the obstacle and dissipation loss due to absorption within the obstacle. In the absence of the obstacle T_p is unity, since then $P_i = P_r$ as a result of the matched termination, and all other quantities in equation (3) are zero. In each experiment, therefore, despite the differing procedures already mentioned, the right-hand side of equation (3) can be evaluated since s and T_p are measured and P_{ad}/P_i can be calculated from the ratio of the crystal currents with and without the discharge tube present. The results are shown graphically in Figure 3 where the ratio P_{ad}/P_i is plotted as a function of the discharge-tube current I_D .

The close agreement between the three experiments is now seen: only at the high values of I_D is there any appreciable separation of the individual curves where the data of Experiment 1 appear to be low compared with the other data. This experiment was less sensitive than the other two, owing to the large mismatch which was introduced when the tube was inserted initially and which remained throughout to mask the much smaller variations due to the discharge itself. The particularly close agreement between the data of Experiment 2 and the average data should also be noted because, for further calculation, the quantities T_p and τ_i are required, where these parameters are measured relative to the zero-current case, i.e. in the form of the data of Experiment 2. Now equation (3) shows that T_p and τ_i cannot both be obtained from the average values of P_{ad}/P_i ; however, if the values of τ_i obtained from Experiment 2 are used, then values of T_p may be calculated from the average data curve of Figure 3. The resulting smoothed data are presented in Table 1.

TABLE 1

I_D (A.)	..	0.00	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.60
τ_i	..	0.005	0.014	0.016	0.022	0.021	0.014	0.013	0.011	0.045	0.062	0.184
T_p	..	0.995	0.87	0.79	0.70	0.615	0.53	0.43	0.32	0.205	0.15	0.14

(b) Determination of Discharge Conductivity

It is shown in Appendix I that the parameters T_p and τ_i are related to the propagation constant γ_2 of the discharge or dispersive medium and also to the thickness d of medium traversed by the radiation. In the analysis a rectangular slab is assumed, which is not true of the experimental arrangement; however, if the thickness of the slab d is taken to be the average thickness of discharge over the crossover section, the loss of accuracy is not too great and the further analysis is facilitated.

By virtue of the fact that the medium is lossy, γ_2 is a complex function of the real and imaginary components of the conductivity of the medium (denoted by σ_r and σ_i respectively). In consequence, the evaluation of σ_r and σ_i from the exact equations, knowing T_p and τ_i , is a difficult process and previous workers in the field have made some form of approximation (usually $d \ll \lambda$) in order

to simplify the calculations. This approximation is indeed far from true in the present instance, since $d=3.25$ cm., which is comparable with the wavelength λ , and an alternative approach is desirable. This approach is to be found in the assumption, $\gamma_2 d$ large, which also yields some simplification of the exact equations and is a condition fulfilled by a dense discharge in which the absorption is high, as in the present case. With this condition imposed, it may be shown (see Appendix I) that

$$T_p = 16e^{-2\alpha_2 d} \left[\frac{\beta_1^2(\alpha_2^2 + \beta_2^2)}{\{\alpha_2^2 + (\beta_1 + \beta_2)^2\}^2} \right], \quad \dots\dots\dots (4)$$

and

$$\eta = \frac{\alpha_2^2 + (\beta_1 - \beta_2)^2}{\alpha_2^2 + (\beta_1 + \beta_2)^2}, \quad \dots\dots\dots (5)$$

where

$$\gamma_2 d = (\alpha_2 + j\beta_2)d \text{ and is large,}$$

$$\beta_1 = 2\pi/\lambda_g, \quad (\lambda_g \text{ is the guide wavelength})$$

and α_2 and β_2 are related to σ_r and σ_i through the expressions:

$$\alpha_2 \beta_2 = \frac{1}{2} \omega \mu_0 \sigma_r, \quad \dots\dots\dots (6)$$

$$\alpha_2^2 - \beta_2^2 = -[(2\pi/\lambda_g)^2 + \omega \mu_0 \sigma_i], \quad \dots\dots\dots (7)$$

where ω is the angular frequency of the radiation and μ_0 the permeability of free space.

TABLE 2

I_D (A.)	σ_r (mho/m.)	σ_i (mho/m.)	σ_i/σ_r
0.10	0.017	-0.109	6.43
0.15	0.022	-0.112	5.08
0.20	0.027	-0.124	4.59
0.25	0.033	-0.122	3.69
0.30	0.040	-0.109	2.74
0.35	0.048	-0.106	2.21
0.40	0.063	-0.090	1.43
0.45	0.068	-0.151	2.22
0.50	0.078	-0.165	2.12
0.60	0.124	-0.225	1.81

The method that has been used and which is considered to be the most practical to determine σ_r and σ_i from the experimental values of T_p and η is as follows: Values of T_p and η are computed from equations (4)-(7) for a range of arbitrarily chosen values of σ_r and σ_i which satisfy the imposed condition $\gamma_2 d$ large, via equations (6) and (7). These computations enable two sets of curves to be drawn, the first set or family relating T_p , σ_r , and σ_i , and the second

family relating η , σ_r , and σ_i . On each set the respective experimental values of T_p or η can also be introduced and the corresponding values of σ_r and σ_i then deduced by interpolation.

The results of this method of evaluation are tabulated in Table 2 and shown graphically in Figure 4 as a function of I_D . The physical significance of the negative sign preceding the values for σ_i is only to account for the inertial

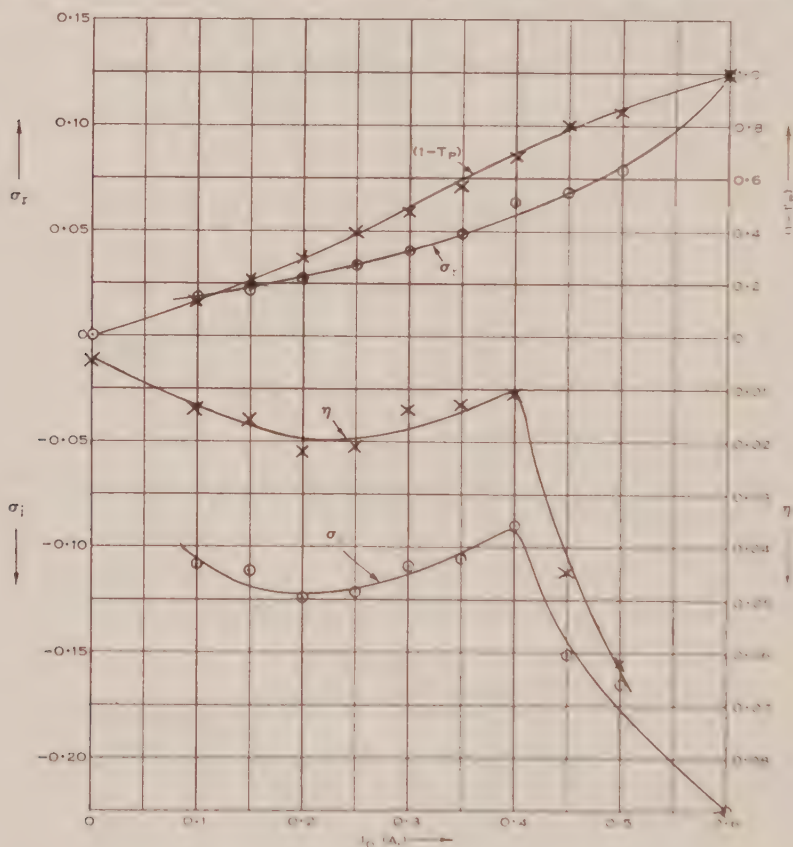


Fig. 4.— σ_r , σ_i , $(1-T_p)$, and η against I_D .

lagging of the electrons in the applied field. Curves of $(1-T_p)$ and of η as functions of I_p are also included in Figure 1. It is interesting to note the resemblance of the σ_r curve to the $(1-T_p)$ curve, and similarly that of the σ_i curve to the η curve.

IV. DISCUSSION

The interaction of electromagnetic radiation with ionized gases has been studied theoretically and experimentally over a period of many years. The techniques used have ranged from the quasi-optical methods of the early workers (Danzer 1929 ; Szelekey 1929 ; Hasselbeck 1932) through a "lumped-constant" era (Appleton and Childs 1930 ; Appleton and Chapman 1932 ; Darrow 1932) to the application of modern microwave methods (Margenau *et al.* 1946 ; Kerr,

Brown, and Kerr 1947; Andrew, Axford, and Sugden 1948; Goldstein and Cohen 1948; Adler 1949; Hawkins and Costain 1949; Rose *et al.* 1949; Rosen 1949; Denno, Prime, and Craggs 1950; Denno 1951). The aim has undoubtedly been to produce experimental systems amenable to theoretical treatment and yet which did not disturb the discharge. The report by Rose *et al.* (1949) reviews many microwave techniques and points out that for high current density discharges, low- Q devices are desirable. Further, in cases where the discharge region is strongly coupled to the transmission line or waveguide, they suggest the use of a four-terminal network method. In this respect, therefore, the technique described in the present paper is entirely consistent with the above observations, since the section of guide containing the discharge is essentially a four-terminal network and the Q of the system is low. Margenau *et al.* (1946) describe a technique very similar to that used by the writer, but the associated measurement of η was not made in their case.

Few of the papers referred to discuss the variation of absorption (or conductivity) as a function of discharge current. Goldstein and Cohen (1948) show a curve of attenuation against current for a discharge in argon (5 mm. pressure) at 2000 Mc/s. which exhibits a maximum with increase in current. Beyond this crest value the attenuation decreases rapidly. They relate this effect to the production of plasma oscillations within the discharge; no corresponding effect has been noted in the present work.

In many of the pre-war publications the results are discussed in terms of either the dielectric constant ϵ and the conductivity σ , or the refractive index n and the extinction or absorption coefficient K . These parameters are related to those previously defined through the expressions,

$$\epsilon = \epsilon_r - j\epsilon_i = (\epsilon_0 + \sigma_i/\omega) - j\sigma_r/\omega,$$

and

$$n^2 - K^2 = \epsilon/\epsilon_0, \quad 2nK = \sigma/\epsilon_0\omega,$$

for a non-magnetic medium.

Hasselbeck (1932, Fig. 6) shows an exact proportionality between current and the absorption coefficient for a neon-helium mixture at a pressure less than 0.1 mm. Hg. At higher pressures the proportionality no longer exists and the curve becomes concave upwards (*loc. cit.*, Fig. 7). This feature is, of course, reproduced in Figure 4 of the present paper, where the curve relating σ_r and I_D is also concave upwards; the correlation between σ_r and K is seen from the previous expressions.

Adler (1949) describes the determination of σ_r and σ_i at a frequency of 8460 Mc/s., using a D.C. sustained mercury-vapour discharge and a cavity-resonator technique. Adler's results show that the absorption is a linear function of current over the range 0–0.8 mA., and that the ratio σ_i/σ_r , corresponding to a gas pressure of 10 mm. Hg, increases with decrease in current. These facts agree with the author's findings; however, a difference exists in that Adler finds the ratio σ_i/σ_r to be less than 1, whereas the values of Table 2 are all greater than 1.

Attention must now be drawn to certain features of the results that, as far as is known, do not appear to have been discussed elsewhere. The similarity of the curve of σ_r against I_D to that of $(1 - T_p)$ against I_D has already been pointed out, the reason for this effect being that the dominant factor in expression (4) for T_p is $e^{-2\alpha_2 d}$, and, from equation (6), α_2 is directly proportional to σ_r . The corresponding similarity between the curve of σ against I_D to that of γ_i against I_D has also been pointed out, and a close inspection of the family of curves relating γ_i , σ_r , and σ , which are used in the analysis, shows that, within the range $0.1 < \sigma_i < 0.3$ mho/m., $(\partial\gamma_i/\partial\sigma_r)\sigma_r$ is very high, whilst $(\partial\gamma_i/\partial\sigma_r)\sigma_i$ is correspondingly small: this is then the reason for the similarity between the curves of γ_i and σ , against I_D because it is precisely this range of σ_i in which the experimental values fall.

A second feature of the curves of γ and σ_i against I_D , namely, the turning point at a current of the order of 0.4 A., signifies that at this value of current the ionized gas is becoming so dense that almost complete reflection occurs at the incident face of the discharge tube. It was observed that the luminescent column spread to the walls of the glass tube at approximately $I_D = 0.4$ A., whereas for smaller currents the visible region did not completely fill the tube. Thus one might expect an increase in current density for $I_D = 0.4$ A., and, in consequence, an increase in γ_i , σ_r , and σ , and a decrease in T_p , as actually observed. It is also possible that these effects may be related to a rapid increase in the scattering coefficient of the discharge at a critical value of the current, as observed by Denno, Prime, and Craggs (1950), Denno (1951), and Rommel (1951), although the increase tended to take place in this case at a current of the order of 1 A., when the same discharge tube was used.

V. CONCLUSIONS

Firstly, the method described employing a low- Q system represents an extension of microwave techniques to higher pressure discharges than have been considered hitherto, although a recent paper by Cobine, Cleary, and Gray (1950) shows a similar trend. Secondly, the volume of discharge involved in the interaction is large, and it may well be that this is a desirable feature because the difficulties associated with boundary and edge effects are probably less significant.

The results show that both the real and imaginary components increase with the discharge current, the real component almost linearly, and the imaginary component, which is negative, non linearly and less rapidly, the curve of σ_i against I_D being concave downward. A further observation is that the ratio σ_r/σ_i tends to decrease with increase in current. These three experimental facts agree with similar published data, but one additional observation can be made, namely, that the ratio σ_r/σ_i is of the order of unity. It has been suggested in a number of theoretical papers that, at microwave frequencies, $\sigma_r \approx \sigma_i$ (since at low and high frequencies $\sigma_r \rightarrow 0$, whereas σ_i tends to a limiting value). It is apparent that for the experimental conditions obtaining here (and also for Adler's work) this proposal breaks down; because the pressure of the gas is

sufficiently high in these cases to make σ_r a vital parameter comparable with σ_i . This consideration does not appear to have been discussed by previous workers in the field.

VI. ACKNOWLEDGMENTS

The author's thanks are due to Professor J. M. Meek, Electrical Engineering Department, University of Liverpool, in whose Department the work was undertaken, and to Dr. J. D. Craggs of the same Department, both of whom showed continual interest in the work during its progress. Thanks are also due to Professor L. G. H. Huxley of the University of Adelaide for valuable comments on the paper during its preparation.

VII. REFERENCES

- ADLER, F. P. (1949).—*J. Appl. Phys.* **20** : 1125.
 ANDREW, E. R., AXFORD, D. W. E., and SUGDEN, T. M. (1948).—*Trans. Faraday Soc.* **44** : 427.
 APPLETON, E. V., and CHAPMAN, F. W. (1932).—*Proc. Phys. Soc. Lond.* **44** : 246.
 APPLETON, E. V., and CHILDS, E. C. (1930).—*Phil. Mag.* **10** : 969.
 COBINE, J. D., CLEARY, E. P., and GRAY, W. C. (1950).—*J. Appl. Phys.* **21** : 1264.
 DANZER, H. (1929).—*Ann. Phys. Lpz.* **2** : 27.
 DARROW, K. K. (1932).—*Bell Syst. Tech. J.* **11** : 576.
 DENNO, S. N. (1951).—Ph.D. Thesis, Univ. Liverpool.
 DENNO, S. N., PRIME, H. A., and CRAGGS, J. D. (1950).—*Proc. Phys. Soc. Lond.* B **63** : 726.
 GOLDSTEIN, L., and COHEN, N. L. (1948).—*Phys. Rev.* **73** : 83.
 HASSELBECK, W. (1932).—*Ann. Phys. Lpz.* **12** : 477.
 HAWKINS, P. O., and COSTAIN, C. C. (1949).—*Nature* **164** : 356.
 KERR, D. E., BROWN, S. C., and KERR, W. P. (1947).—*Bull. Amer. Phys. Soc.* **22** : 28.
 MARGENAU, H., McMILLAN, F. L., DEARNLEY, I. H., PEARSALL, C. H., and MONTGOMERY, C. G. (1946).—*Phys. Rev.* **70** : 349.
 ROMMEL, D. (1951).—*Nature* **167** : 243.
 ROSE, D. J., KERR, D. E., BIONDI, M. A., EVERHART, E. C., and BROWN, S. C. (1949).—M.I.T. Radiation Laboratory Tech. Rep. No. 140.
 ROSEN, P. (1949).—*J. Appl. Phys.* **20** : 868.
 SMULLIN, L. D., and MONTGOMERY, C. G. (1948).—"Microwave Duplexers." p. 185. M.I.T. Radiation Laboratory Series, Vol. 14. (McGraw-Hill: New York.)
 STRATTON, J. A. (1941).—"Electromagnetic Theory." p. 511. (McGraw-Hill: New York.)
 SZEKELY, A. (1929).—*Ann. Phys. Paris* **3** : 112.

APPENDIX I

It is required to establish expressions for the power transmission and reflection coefficients of an electromagnetic wave incident on a lossy medium in terms of the physical constants of that medium. This problem has been considered previously in so far as the transmission coefficient only is concerned, and expressions are given by Margenau *et al.* (1946, equation (12)) and by Smullin and Montgomery (1948, equation (64)). The latter cite Stratton (1941) for the derivation of their expression. There are, unfortunately, certain differences between the expressions given in the above paper and textbooks, but an examination of Stratton's derivation shows that the differences are due to the form of boundary conditions assumed by the various workers.

(a) Derivation of T_p and η

Consider the system shown in Figure 5, where a slab of lossy medium, characterized by the propagation constant γ_2 , is located between two non-lossy homogeneous media characterized by the propagation constants γ_1 and γ_3 . Let the thickness of the slab be d and let the axis of propagation of an electromagnetic wave be normal to the slab. Define the amplitudes of the component forward and backward waves in each of the three media as follows:

- (i) medium 1 (γ_1) $E_{1i}, E_{1r}, H_{1i}, H_{1r}$, measured at $z=0$,
- (ii) medium 2 (γ_2) $E_{2i}, E_{2r}, H_{2i}, H_{2r}$, measured at $z=0$,
- (iii) medium 3 (γ_3) E_{3i}, H_{3i} , measured at $z=d$.

(It is assumed that medium 3 is correctly terminated so that there is no backward wave.)

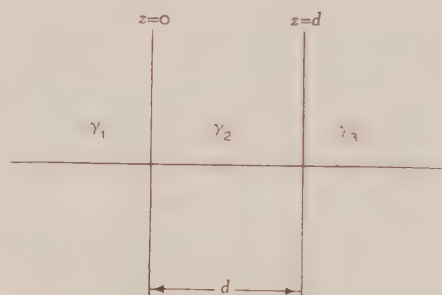


Fig. 5

At the first interface ($z=0$) the boundary conditions are

$$E_{1i} + E_{1r} = E_{2i} + E_{2r}, \dots\dots\dots (\text{A1})$$

$$H_{1i} + H_{1r} = H_{2i} + H_{2r}, \dots\dots\dots (\text{A2})$$

and the boundary conditions at the second interface are

$$E_{2i}e^{-\gamma_2 d} + E_{2r}e^{+\gamma_2 d} = E_{3i}, \dots\dots\dots (\text{A3})$$

$$H_{2i}e^{-\gamma_2 d} + H_{2r}e^{+\gamma_2 d} = H_{3i}, \dots\dots\dots (\text{A4})$$

The electric and magnetic field vectors are related through the intrinsic impedance of the medium according to the relations

$$\frac{E_{1i}}{H_{1i}} = -\frac{E_{1r}}{H_{1r}} = Z_1, \quad \frac{E_{2i}}{H_{2r}} = -\frac{E_{2r}}{H_{2i}} = Z_2, \quad \frac{E_{3i}}{H_{3i}} = Z_3.$$

Solving the above equations for the ratio E_i/E_r , which is defined as the transmission coefficient of the slab and denoted T , we find

$$T = \frac{-4Z_2Z_1}{[e^{-\gamma_2 d}(Z_1 - Z_2)(Z_3 - Z_2) - e^{+\gamma_2 d}(Z_1 + Z_2)(Z_3 + Z_2)]}, \dots (\text{A5})$$

Assuming media 1 and 3 are the same by setting $Z_1 = Z_3$ and $\gamma_1 = \gamma_3$, equation (A5) reduces to

$$T = \left[\cosh \gamma_2 d + \frac{1}{2} \left(\frac{Z_1 + Z_2}{Z_2} + \frac{Z_2}{Z_1} \right) \sinh \gamma_2 d \right]^{-1}, \dots\dots\dots (\text{A6})$$

This result is essentially similar to that given by Margenau *et al.* (1946) except that it is hyperbolic in form rather than trigonometrical. "Removing" medium 2 by setting $\gamma_1 = \gamma_2$ and $Z_1 = Z_2$ yields the expected result

$$T = e^{-\gamma_1 d} = e^{-j\beta_1 d},$$

since α_1 is assumed zero for medium 1.

Now the intrinsic impedance of the medium is defined in terms of the constants of the medium by the equation

$$Z_k = \frac{-j\omega\mu_k}{\gamma_k} \dots\dots\dots (\text{A7})$$

Assuming for the purpose of the analysis that we can set $\mu_1 = \mu_2 = \mu_0$ (the permeability of free space), equation (A6) becomes

$$T = \left[\cosh \gamma_2 d + \frac{1}{2} \left(\frac{\gamma_2}{\gamma_1} + \frac{\gamma_1}{\gamma_2} \right) \sinh \gamma_2 d \right]^{-1} \dots\dots\dots (\text{A8})$$

The analysis given thus far is not impaired in any way if we regard the system to be located not in free space but within a waveguide operating in the *H*-dominant mode, because the boundary conditions remain unaltered. However, any further analysis must observe the location of the system because the propagation constant within a waveguide is not the same as in free space. For a lossy medium, as considered in the present instance, the propagation constant within the waveguide is defined by the relation

$$(\gamma_2)^2 = \left(\frac{\pi}{a} \right)^2 - \left(\frac{2\pi}{\lambda} \right)^2 + j\omega\mu_0(\sigma_r + j\sigma_i), \dots\dots\dots (\text{A9})$$

where λ is the "free space" wavelength of the radiation and $2a$ is the dimension of the broad face of the guide.

Similarly, for the non-lossy medium, the propagation constant is defined by

$$(\gamma_1)^2 = - \left(\frac{2\pi}{\lambda_g} \right)^2, \dots\dots\dots (\text{A10})$$

where λ_g is the guide wavelength in medium 1.

Substitution in equation (A8) for γ_1 and γ_2 from equations (A9) and (A10) does not lead to any simple result owing to the complex form of γ_2 .

The parameter of practical significance, however, is not the coefficient T but the associated quantity T_p , the power-transmission coefficient, where

$$T_p = \frac{E_{3i}^2}{|E_{1i}|^2} = |T|^2 \dots\dots\dots (\text{A11})$$

An expression for the reflection coefficient of the slab can be established readily in terms of the voltage standing-wave ratio in medium 1 as follows:

Within medium 1 the voltage standing wave is defined by the equation

$$E_1 = E_{1i} e^{-\gamma_1 z} + E_{1r} e^{+\gamma_1 z},$$

where E_1 is the resulting field amplitude at distance z (negative) from the origin, $z=0$.

From normal transmission-line theory, the voltage standing-wave ratio s in this medium is defined by the equation

$$s = \frac{E_{1max}}{E_{1min}} = \frac{1 + |r_1|}{1 - |r_1|},$$

or

$$|r_1| = (s-1)/(s+1), \quad \text{..... (A12)}$$

where $r_1 = E_{1r}/E_{1i}$ and is the complex reflection coefficient of the interface.

Now from the boundary conditions we obtain the relation

$$E_{1i} \left(1 + \frac{Z_2}{Z_1}\right) + E_{1r} \left(1 - \frac{Z_2}{Z_1}\right) = E_{3i} e^{\gamma_2 d} \left(1 + \frac{Z_2}{Z_3}\right),$$

and dividing throughout by E_{1i} yields

$$\left(1 + \frac{Z_2}{Z_1}\right) + r_1 \left(1 - \frac{Z_2}{Z_1}\right) = T e^{\gamma_2 d} \left(1 + \frac{Z_2}{Z_3}\right), \quad \text{..... (A13)}$$

where T is the transmission coefficient of the slab.

Assuming media 1 and 3 to be similar, equation (A13) reduces to

$$r_1 = \frac{Z_1 + Z_2}{Z_1 - Z_2} (T e^{\gamma_2 d} - 1). \quad \text{..... (A14)}$$

Substituting for Z_1 and Z_2 from equation (A7) and combining with (A12), again assuming that $\mu_1 = \mu_2 = \mu_0$, equation (A14) becomes

$$\eta = \frac{|\gamma_2 + \gamma_1|}{\gamma_2 - \gamma_1} (T e^{\gamma_2 d} - 1)^{-2}. \quad \text{..... (A15)}$$

(b) Approximate Solutions for Small $\gamma_2 d$

Both Margenau *et al.* (1946) and Smullin and Montgomery (1948) derive approximate solutions for the transmission coefficient subject to the condition, $\gamma_2 d$ small. This assumption removes the difficulties inherent in the solution of the exact equations.

Noting that $\cosh \gamma_2 d \rightarrow 1$ and $\sinh \gamma_2 d \rightarrow \gamma_2 d$ for small $\gamma_2 d$, equation (A8) simplifies to

$$T = [1 + \frac{1}{2} d \gamma_1 (1 + \gamma_2^2 / \gamma_1^2)]^{-1}. \quad \text{..... (A16)}$$

Introducing two parameters,

$$x = \frac{1}{2} \omega \mu_0 d (\lambda_g / 2\pi) \quad \text{and} \quad y = d (2\pi / \lambda_g),$$

and combining equations (A16), (A9), and (A10), we obtain

$$T_p = [(1 + x\sigma_r)^2 + (y + x\sigma_i)^2]^{-1}. \quad \text{..... (A17)}$$

The expressions defining the parameters x and y show that, for high frequencies and sufficiently small d , $x \gg y$ and equation (A17) may be written more simply

$$T_p = [(1 + x\sigma_r)^2 + (x\sigma_i)^2]^{-1}, \quad \text{..... (A18)}$$

which is the limiting form given by both Margenau *et al.* (loc. cit., p. 355) and Smullin and Montgomery (loc. cit., equation (68)).

The same approximation can be applied to equation (A15) by substituting for T in equation (A16) and noting that, for small $\gamma_2 d$, $e^{\gamma_2 d} \rightarrow 1 + \gamma_2 d$, whence it may be shown that

$$\eta = (\sigma_r^2 + \sigma_i^2) / (\sigma_r + 1/x)^2 + \sigma_i^2. \quad \dots\dots\dots (\text{A19})$$

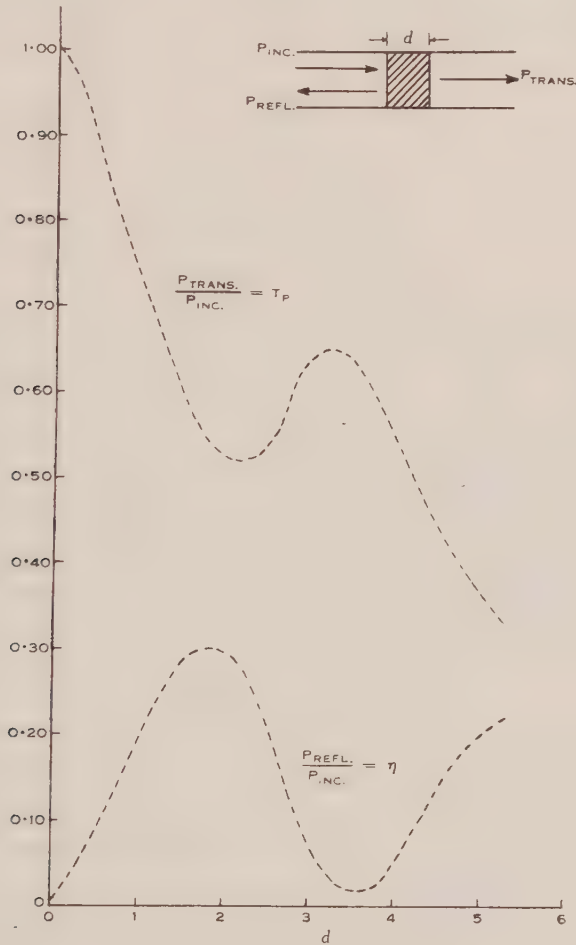


Fig. 6.— T_p and η against d .

(c) Approximate Solutions for Large $\gamma_2 d$

Noting that for large $\gamma_2 d$ (for 2 per cent. accuracy, $\gamma_2 d > 2$)

$$\cosh \gamma_2 d = \sinh \gamma_2 d = e^{\gamma_2 d} / 2 ;$$

then equation (A8) reduces to

$$T \simeq \left[\frac{e^{\gamma_2 d}}{4} \frac{(\gamma_1 + \gamma_2)^2}{\gamma_1 \gamma_2} \right]^{-1}, \quad \dots\dots\dots (\text{A20})$$

or, in terms of $T_p (=|T|^2)$, α_2 , β_2 , and β_1 ,

$$T_p = 16e^{-2\alpha_2 d} \left[\frac{\beta_1^2(\alpha_2^2 + \beta_2^2)}{\{\alpha_2^2 + (\beta_1 + \beta_2)^2\}^2} \right], \dots\dots\dots (\text{A21})$$

where $\alpha_2 + j\beta_2 = \gamma_2$ and $j\beta_1 = \gamma_1$ (see also equations (6) and (7), Section III (b)).

In the case of the reflection coefficient, the same approximation provides the following simplification. Starting from equation (A15) and substituting for T from equation (A20) yields

$$\tau_i = \left| \frac{\gamma_1 - \gamma_2}{\gamma_1 + \gamma_2} \right|^2, \dots\dots\dots (\text{A22})$$

or

$$\tau_i = \frac{\alpha_2^2 + (\beta_1 - \beta_2)^2}{\alpha_2^2 + (\beta_1 + \beta_2)^2}, \dots\dots\dots (\text{A23})$$

(d) Discussion of Analysis

Equations (A8) and (A15) are of little more than academic interest, particularly when the lossy medium is such that γ_2 is complex, owing to the inherent difficulty of solution. However, they show that both the transmission coefficient T and the reflection coefficient τ_i are oscillating functions of the thickness d and that these functions will be exponentially damped as indicated in Figure 6. The physical significance of the two forms of approximation, $\gamma_2 d$ small and $\gamma_2 d$ large, may now be appreciated. In the former case, the analysis is restricted to that range of d corresponding to the first half-cycle of the oscillating functions, whereas for the latter case, the analysis is restricted to the range of d for which the oscillations have died out.

In regard to the assumption, $\gamma_2 d$ large, it has already been stated that

$$\cosh \gamma_2 d = \sinh \gamma_2 d = e^{\gamma_2 d/2},$$

with 2 per cent. accuracy for $\gamma_2 d > 2$.

Inserting the experimental value of d (3.25×10^{-2} m.), the corresponding requirement is (approximately) $\gamma_2 = 60$. In practice, however, γ_2 is complex and it is therefore necessary to consider the condition in terms of the real parameters, α_2 and β_2 , rather than γ_2 .

For $\alpha_2 = 60$, it is apparent that the value of β_2 does not affect the accuracy of the assumption and the values of α_2 involved in the numerical analysis show that this requirement is met.

THE MICROWAVE ADMITTANCE OF A MERCURY-VAPOUR DISCHARGE

By H. A. PRIME*

[*Manuscript received March 3, 1952*]

Summary

The normalized conductance and susceptance of a low-pressure mercury-vapour discharge have been measured as a function of the discharge current at a frequency of c. 10,000 Mc/s. by a microwave method. The complex conductivity of the discharge is then calculated from the admittance data by representing the discharge as a shunt admittance loading the waveguide. The validity of this approximate representation is examined. The conductivity results are in agreement with similar published results and the similarity between the characteristics of a D.C. sustained discharge and a radio-frequency discharge are noted. It is deduced that the collision frequency varies only slowly with change in discharge current.

I. INTRODUCTION

The experimental technique described in this paper is based on a four-terminal network method and provides for the measurement of the normalized conductance and susceptance of a mercury-vapour discharge as a function of the discharge current. The analysis of the data so obtained to determine the complex conductivity of the discharge raises a number of electromagnetic problems concerning the representation of the discharge as a single shunt admittance loading the waveguide.

Previous work on this subject is extremely limited and only two papers containing experimental data on admittance are known to the author (Everhart and Brown 1949; Mumford 1949). Their results are compared with those reported in this paper and a number of interesting features are developed, for example, the similarity between the admittance characteristics of a radio-frequency sustained discharge and a D.C. sustained discharge.

The point must also be made, as in a previous paper (Prime 1952), that the microwave power level used for the measurements is inadequate to influence the discharge and does not contribute to the ionization processes in any way.

II. EXPERIMENTAL APPARATUS AND METHOD

A schematic diagram of the experimental system is shown in Figure 1 and the design of the discharge-tube section is shown in Figure 2.

The mercury-vapour discharge used for this work was provided by a commercial ultraviolet discharge lamp manufactured by Hanovia Ltd., Slough, England. The internal operating pressure of the tube was of the order of 10 mm. Hg at normal room temperature (20 °C.) and the radiation emitted by the

* Electrical Engineering Department, University of Adelaide.

discharge was predominantly located at the 2537 Å resonance line of mercury. The tube also contained a little argon to assist the initiation of the discharge, since it was of the cold-cathode type, operated from a 3000 V. D.C. supply. The current through the discharge was controlled by varying a current-limiting resistor in series with the supply.

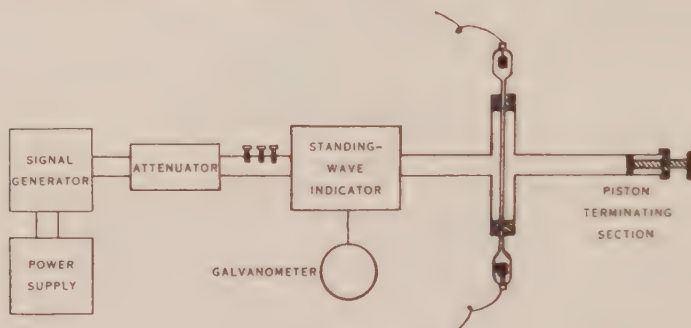


Fig. 1.—Block diagram of apparatus.

The series cylinders on each side of the guide (Fig. 2) were necessary because the diameter of each end of the discharge tube was approximately $\frac{1}{8}$ in. (to contain the electrode assembly) and a $\frac{1}{2}$ in. hole was required through the waveguide

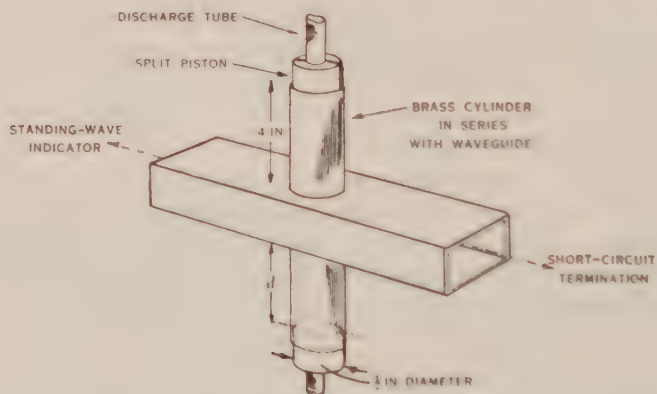


Fig. 2.—Discharge-tube section.

to permit the insertion of the tube. Unfortunately, this $\frac{1}{8}$ in. diameter requirement did not create an evanescent-mode condition in the two cylinders and, as normal mode propagation was possible in these arms, split pistons were introduced and adjusted to produce a virtual short circuit across the holes in the waveguide at the positions of entry and exit of the tube. In fact it was found that the adjustment of the pistons was not critical, thus showing that little power escaped into the cylinders from the main guide.

The preliminary adjustments of the system, necessary before measurements could be made, were as follows: Without the discharge tube in position, the

setting of the coarse attenuator was increased until variation of the terminating-piston position did not react on the power output of the klystron oscillator, thus showing that the klystron was adequately decoupled from the main waveguide run. The discharge tube was then introduced, the tube supply switched on, and the discharge current adjusted to some convenient value about the middle of the possible range of variation (8–30 mA.). The position of the terminating piston was then adjusted to give a standing-wave ratio of the order of 1.2 or 1.3 at the mid-value of current. This adjustment ensured that the system did not become seriously mismatched and yield a high standing-wave ratio when the current was varied between the limiting values.

Four similar experiments were carried out for different initial settings of the terminating-piston position. In each experiment, the current through the discharge tube was varied over the range, 8–30 mA., and, corresponding to each value of current selected, the standing-wave ratio and the distance from the first minimum of the standing wave to the tube axis were noted. In addition, the positions of successive maxima and minima of the standing wave were noted on the indicator scale in order to permit subsequent calculation of the guide wavelength and also for comparison with the same quantity derived from the cavity-wavemeter frequency reading.

III. EXPERIMENTAL RESULTS

(a) *Measurement of Total Admittance*

The frequency maintained throughout the experiments and measured on the cavity wavemeter was 9300 Mc/s., corresponding to a guide wavelength λ_g of 4.21 cm. The mean value of λ_g measured on the standing-wave indicator from the maxima and minima of the standing-wave pattern was 4.13 cm. The overall mean value was therefore taken to be 4.17 cm., the accuracy being of the order of 2 per cent.

The measurements of the standing-wave ratio s and distance to the first minimum of the standing-wave pattern from the tube axis l enabled the total admittance of the system y_T to be calculated from the expression

$$y_T = \frac{s + j \cot 2\pi x}{1 + js \cot 2\pi x}, \quad \dots \dots \dots (1)$$

where y_T is the normalized total admittance in a transverse plane through the axis of the discharge tube and $x = l/\lambda_g$. Alternatively, the calculation may be more readily effected by use of a circle diagram, this being the method used in practice. The results of the experiments are tabulated in Table 1, where $y_T = g_T - jb_T$ and I_D is the discharge-tube current. The overall accuracy of the measurements is estimated to be about 5 per cent.

The radius of the discharge (i.e. internal bore of tube) was also measured and found to be 1 mm.

(b) *Determination of Discharge Admittance*

By defining the distance l to be from the axis of the discharge tube to the position of the first minimum on the standing-wave indicator scale, the value of

y_T , as stated, represents the admittance of the system referred to a plane containing the tube axis and normal to the guide axis. Consideration must now

TABLE 1

Experiment 1			Experiment 2			Experiment 3			Experiment 4		
I_D (mA.)	g_T	b_T	I_D (mA.)	g_T	b_T	I_D (mA.)	g_T	b_T	I_D (mA.)	g_T	b_T
0.0	0.875	-0.120	0.0	0.793	-0.140	0.0	0.765	-0.063	0.0	0.745	-0.077
9.4	1.000	-0.218	9.6	0.895	-0.195	9.2	0.895	-0.190	8.7	0.850	-0.180
17.2	1.045	-0.245	18.2	0.975	-0.215	16.0	0.945	-0.220	15.0	0.885	-0.195
23.0	1.066	-0.250	23.5	1.020	-0.240	23.3	0.995	-0.245	23.0	0.930	-0.225
29.5	1.090	-0.275	29.5	1.065	-0.270	29.7	1.030	-0.275	29.7	0.970	-0.265

be given to the various component admittances which go to make up the total admittance y_T . These may be listed as follows:

- (i) The admittance y_{sc} of the short-circuit termination referred to the tube axis. Theoretically, if the short circuit is ideal, then y_{sc} is a pure susceptance and is dependent only on the distance from the tube axis to piston face.
- (ii) The admittance y_w of the waveguide region containing the side cylinders. This is admittedly a somewhat indeterminate admittance and its contribution and representation are therefore difficult to assess.
- (iii) The admittance y_t of the discharge tube itself (no discharge present). Again, from theoretical considerations, this component should be mainly susceptive, since the wall of the tube is quartz and losses are therefore probably negligible.
- (iv) The admittance y_D of the discharge within the tube. This is the component in which we are ultimately interested for the purpose of further analysis.

It is now *assumed* that the total admittance of the system is made up from the parallel combination of the above components, so that

$$y_T = y_{sc} + y_w + y_t + y_D \quad \dots \dots \dots (2)$$

The inherent difficulty of the assumption is that it is by no means certain that the individual factors can be represented by an equivalent shunt admittance. This consideration is particularly true of the (ii), (iii), and (iv) components, where it is known for example that a *thick*, dielectric post placed parallel to the E -vector across a waveguide cannot be represented by a simple shunt susceptance but is required to be interpreted in terms of a four-terminal T-network (Huxley 1917). However, as the diameter of the discharge tube is of the order of $\lambda_c/8$ and the diameter of the discharge itself is of the order of $\lambda_c/20$, "thick post"

considerations can probably be neglected, thus leaving the y_w component as the one most likely to invalidate the assumption. Now equation (2) can be rewritten

$$y_T = y_{D_0} + y_D, \quad \dots \dots \dots (3)$$

where

$$y_{D_0} = y_w + y_{sc} + y_t, \quad \dots \dots \dots (4)$$

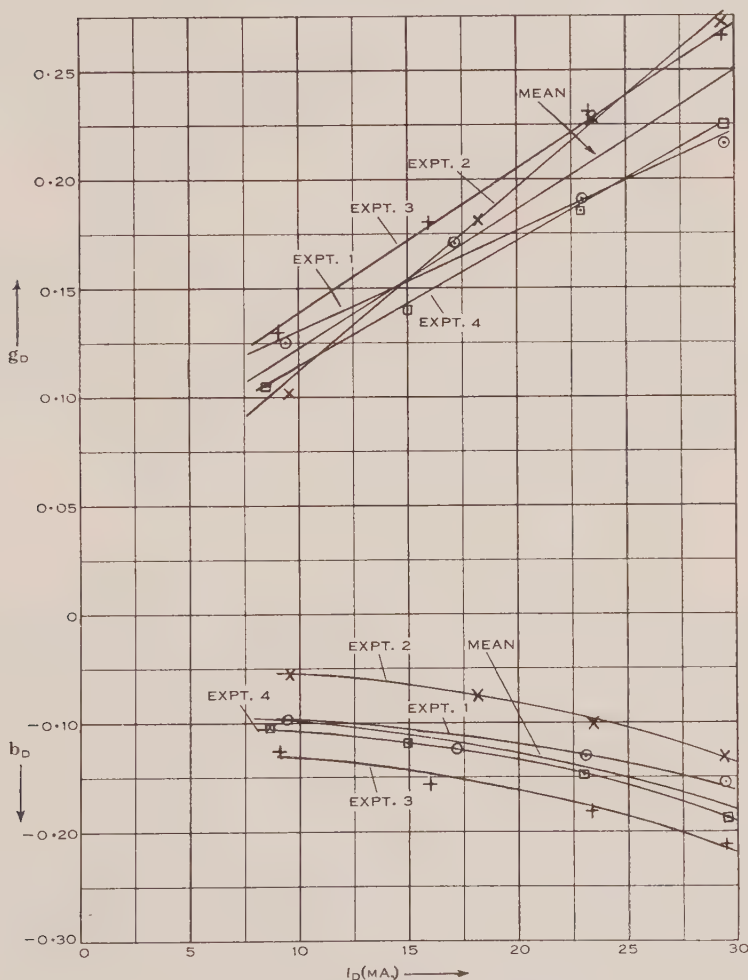


Fig. 3.— y against I_D .

and represents the admittance of the system for zero discharge current. y_D may therefore be calculated as a function of I_D by application of equation (3): the results are shown in graphical form in Figure 3. It will be seen that g_D is a linearly increasing function of I_D for currents greater than approximately 8 mA., whereas b_D , which is negative, increases only slowly and non-linearly with I_D .

(c) *Determination of Discharge Conductivity*

The relationship between y_D and the components σ_r and σ_i of the complex conductivity has been discussed theoretically and demonstrated experimentally by Everhart and Brown (1949). This analysis is concerned primarily with a discharge maintained by a radio-frequency field between parallel plates such that the radius of the discharge r_0 is several times larger than the plate spacing δ ; in addition, $r_0 \ll \lambda$, the wavelength of the radiation. Thus, the boundary-value problem which they solve is quite distinct from that obtaining for the present D.C. sustained discharge and the application of their expressions to the present problem is questionable. However, for the purpose of the subsequent discussion, a single computation has been carried out using their admittance relation:

$$\frac{G_D \delta}{\pi r_0^2 \omega \epsilon_0} + \frac{j B_D \delta}{\pi r_0^2 \omega \epsilon_0} = \delta \left[\omega \epsilon_0 \int_{-\delta/2}^{+\delta/2} \frac{dz}{\sigma_{C_0} \cos(\pi z/\delta) + j \omega \epsilon_0} \right], \quad \dots \quad (5)$$

where δ and r_0 have been previously defined, and ω and ϵ_0 have their usual significance. The axis of integration z is the discharge axis and σ_{C_0} is the complex conductivity of the discharge, equal to $(\sigma_r + j\sigma_i)$, measured at $z=0$. For the purpose of computing σ_{C_0} from the above integral equation, Everhart and Brown include in their paper curves relating G_D , B_D , and σ_{C_0} , and the electron-collision frequency ν_e . Using the average experimental data of the present work, corresponding to a current of 16 mA., it is found that

$$G_D = 3.28 \times 10^{-4} \text{ mho}, \quad B_D = -2.33 \times 10^{-4} \text{ mho},$$

whence

$$|\sigma_{C_0}| = 1.83 \text{ mhos/m. and } \nu_e = 2.92 \times 10^{10} \text{ c/s.}$$

An alternative approach to what is virtually the same problem has been given by Frank (1942). In this analysis, a dielectric post of circular cross section is assumed to extend across the waveguide, such that the axis of the post is parallel to the electric field of the dominant mode. Thus the boundary-value problem so postulated is identical with that of the discharge column across the waveguide used for the present work and greater significance can be attached to Frank's analysis. Frank then finds that, by the method of images, the problem of the behaviour of the system is equivalent to that of an infinite grating composed of parallel dielectric cylinders with a grating spacing of a , the wider dimension of the guide. For the important case in which the post radius r is small compared with the free-space wavelength λ (i.e. $(2\pi r/\lambda)^2 \ll 1$), the post behaves as a lumped-constant admittance shunted across the waveguide. If the normalized value of this admittance be denoted by y , the following expression is given as a good approximation under the above conditions:

$$y = j \frac{a}{2\lambda_g} \left[\ln \lambda/r - f(a/\lambda) - \frac{1}{2K_e} (\lambda/\pi r)^2 - 1.47 \right], \quad \dots \dots \dots (6)$$

where K_e is the complex dielectric constant of the post. The function $f(a/\lambda)$ is shown in graphical form by Frank for the useful range $\frac{1}{2} \leq a/\lambda \leq 1$, over which it is almost linear and satisfies the equation:

$$f(a/\lambda) = -1.78a/\lambda + 2.214. \quad \dots \dots \dots (7)$$

Combining equation (6) with (7) and rearranging yields

$$K_e = \frac{(\lambda/\pi r)^2}{2[\ln \lambda/r + 1.78a/\lambda - 3.684 + j2\lambda_g/ay]} \dots\dots\dots (8)^*$$

Finally, equation (8) can be expressed in terms of the real and imaginary parts of the complex conductivity rather than the complex dielectric constant ;

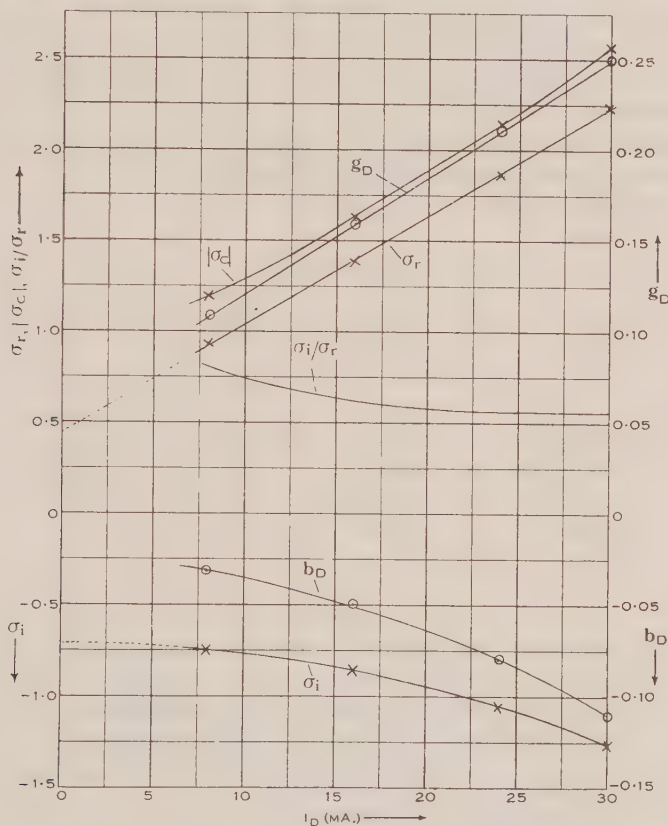


Fig. 4.— σ_r , σ_i , g_D , b_D , $|\sigma_c|$, and σ_i/σ_r against I_D .

with this modification and inserting the numerical constants of the system for λ , r , and a , the equation becomes

$$(1 + \sigma_i/\omega\epsilon_0) - j\sigma_r/\omega\epsilon_0 = \frac{106y_D}{[2.4y_D + j6.6]} \dots\dots\dots (9)$$

Noting that σ_r and σ_i are both zero for $I_D=0$, it follows from (9) that

$$g_D=0, \quad b_D=0.064, \quad \text{for } I_D=0.$$

Since the shunt conductance g_D is due to the electrons which provide a conducting channel across the guide, one expects and finds that g_D is zero at $I_D=0$. The

* Montgomery, Dicke, and Purcell (1948) quote Frank's report and give equation (8) in a similar form but with the constant equal to -1.91 instead of -3.684 . The reason for this difference is not clear, particularly as the coefficient of a/λ is the same.

susceptance, however, is due partly to the fact that the electrons cannot follow the applied field due to inertial lagging, thus creating a reactive-current component, and partly to the displacement current, which is also in quadrature with the applied field and which is present irrespective of the presence of electrons, thereby creating the finite value of b_D when $I_D = 0$. However, the susceptance represented by this particular value of b_D is included in the measurement of y_{D0} , whereas actually it should be retained as part of y_D . Accordingly, the values of y_D shown in Table 1 require to be increased by $j0.064$ before the values of σ_r and σ_i are computed by equation (9). As might be expected, the values of σ_r are modified only slightly, whilst the values of σ_i are approximately 0.5 mho m. less than they otherwise would have been. The revised data obtained from the mean data of Figure 3 yield the results given in Table 2 and shown graphically in Figure 4.

TABLE 2

I_D (mA.)	y_D (mean)	σ_r (mhos/m.)	σ_i (mhos/m.)	σ_i/σ_r	σ_{C_0}
0	$0.00 + j0.064$	0.000	0.00	0.00	0.00
8	$0.11 - j0.032$	0.925	-0.75	0.81	1.19
16	$0.16 - j0.05$	1.38	-0.86	0.62	1.63
24	$0.21 - j0.08$	1.86	-1.06	0.57	2.14
30	$0.25 - j0.11$	2.24	-1.27	0.57	2.57

IV. DISCUSSION

The technique is essentially a low- Q , four-terminal network method, which is the type most suited to high current density discharges that are strongly coupled to the guide (Rose *et al.* 1949). It is undoubtedly a more elegant method than the measurement of the transmission and reflection coefficients described by the author (Prime 1952), since the conditions obtaining at the position of the discharge tube within the guide are not so open to speculation, owing to the use of a smaller diameter tube. However, the technique cannot be applied indiscriminately because the derivation of the conductivity depends on the fulfilment of the condition, $(2\pi r/\lambda)^2 \ll 1$, so that the representation of the discharge as a single lumped constant element is valid (in the present case, $(2\pi r/\lambda)^2 = 0.038$).

The spread of the individual experimental curves shown in Figure 3 is of the order of ± 10 per cent. for y_D and somewhat greater for b_D ; the extreme sensitivity to temperature of the vapour pressure of mercury is the probable explanation of this effect, as the experiments were carried out on different days and at different times during the day. The curves have not been extrapolated below 8 mA., as this was the extinction current of the discharge.

The technique used by Everhart and Brown (1949) is more refined than the one described in this paper, in so far as the discharge region was so designed that all the associated theoretical requirements were adequately met. Two features

of their work are of considerable interest with reference to the present discussion. The experimental variation of the discharge admittance, firstly, with variation of discharge current (equivalent to keeping $\gamma = \omega/v_c$ constant and varying $\eta = |\sigma_{c0}|/\omega\epsilon_0$), and secondly, with variation of gas pressure (η constant, γ varying), was found to be in complete agreement with the variation predicted

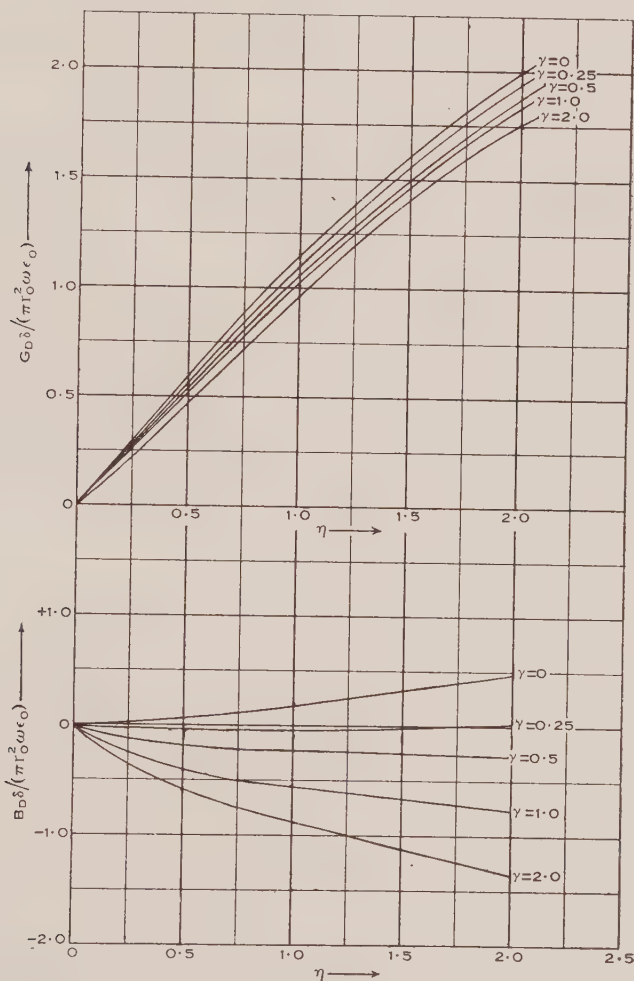


Fig. 5.— G_D and B_D against η , for various values of γ (Everhart and Brown 1949).

by the theoretical analysis. If their curves relating G_D , B_D , γ , and η , are redrawn to show the variation, G_D with η and B_D with η , γ being held constant in each case, the curves of Figure 5 are obtained. The close similarity between these and the corresponding ones of Figure 3 is immediately apparent, particularly for the case of $\gamma=2$. This is significant because the work by Everhart and Brown was undertaken with a radio-frequency sustained helium discharge, whereas the discharge used for the present work was in mercury vapour and

D.C. sustained. It would be expected, therefore, that, in the latter case, γ would not be constant, as I_D varied owing to temperature variations (and hence pressure variations) producing an increase in the collision frequency ν_c as I_D increased, thereby reducing γ . The suggestion is, however, that ν_c is only varying slightly over the current range, 8–30 mA., and that γ can be regarded as a constant to a first approximation. The second feature of interest concerns the calculation of $|\sigma_{c_0}|$ and ν_c using equation (5), when the following values were calculated for $I_D=16$ mA.:

$$|\sigma_{c_0}|=1.83 \text{ mhos/m.}, \quad \nu_c=2.92 \times 10^{10} \text{ c/s.}$$

The analysis of Frank, which was later used, yielded the data of Table 2 from which the corresponding value of $|\sigma_{c_0}|$ is seen to be

$$|\sigma_{c_0}|=1.625 \text{ mhos/m.},$$

which is in remarkably close agreement considering that the two analyses differ so much in approach and concept.

The work of Mumford (1949), on the noise characteristics of a similar mercury-vapour discharge at 4000 Mc/s. (cf. Fig. 4 of his paper) showed that g_n varied linearly with the current and that the susceptance b_n increased negatively up to about 60 mA. and then had a broad maximum over the range 65–100 mA., beyond which it tended to decrease again slowly. Over the lower end of this current range there is, therefore, exact agreement with the form of the curves in Figure 3. In terms of actual values, Mumford finds

$$g_D=0.25, \quad b_D=-0.33, \quad \text{at } I_D=20 \text{ mA.},$$

whereas the corresponding values taken from Figure 3 are

$$g_D=0.19, \quad b_D=-0.13.$$

The somewhat higher values found by Mumford would be expected if the diameter of the discharge tube was greater than the diameter of the one used by the author, which appears to be so from measurements made on the tube assembly figure included in the paper. Greater significance could have been attached to this comparison of actual values had Mumford presented the data in terms of $\sigma_{(\text{complex})}$ rather than of y_D .

In comparing the results with those obtained by other techniques, the work of Adler (1949) is relevant, because, in addition to the frequency and mercury-vapour discharge being common to both experiments, it is apparent that the pressure of gas and dimensions of the discharge must also have been approximately the same in each case.

Unfortunately, Adler's data only cover the current range 70 μ A. – 130 μ A., and it would be dangerous to extrapolate these data by a factor of 100 in order to compare numerical values of $\sigma_{(\text{complex})}$ with those obtained by the author. However, the ratio σ_i/σ , is less than unity in both cases and, as $I_D \rightarrow 0$, the ratio tends towards unity; moreover, there is substantial agreement between the actual numerical values.

V. CONCLUSIONS

The real and imaginary components of the complex conductivity of the discharge have been shown to increase with the discharge current, the real component linearly and the imaginary component, which is negative, non-linearly and less rapidly, the curve being concave downwards. This result confirms previous data appropriate to a mercury discharge (Prime 1952). The further observation that the ratio σ_i/σ_r tends to decrease with increase in current also agrees with similar published data. The application of Frank's analysis for the derivation of the above results therefore appears to be justified.

The similarity between the curves of Y_D against I_D obtained for a radio-frequency sustained discharge by Everhart and Brown and the corresponding curves of the present paper (Fig. 3) has been pointed out, and the observation, that the collision frequency ν_c is probably only varying slowly with variation of current, has also been noted. The similarity might be expected on the grounds that the measurements made in the present case were appropriate to the centre of the discharge region where positive column conditions existed. This portion of a D.C. sustained discharge is known to resemble a radio-frequency sustained discharge in the following ways:

- (a) No net charge—plasma conditions exist,
- (b) Low gas and ion temperatures, but high electron temperatures,
- (c) Voltage gradient less for monatomic than for diatomic gases.

VI. ACKNOWLEDGMENTS

The author's thanks are due to Professor J. M. Meek, Electrical Engineering Department, University of Liverpool, in whose Department the work was undertaken, and to Dr. J. D. Craggs, of the same Department; and to Professor L. G. H. Huxley, of the University of Adelaide, for his kindness and help during the preparation of the paper.

VII. REFERENCES

- ADLER, F. P. (1949).—*J. Appl. Phys.* **20**: 1125.
EVERHART, E., and BROWN, S. C. (1949).—*Phys. Rev.* **76**: 839.
FRANK, N. H. (1942).—M.I.T. Radiation Laboratory Rep. No. T-9, p. 32.
HUXLEY, L. G. H. (1947).—"Wave Guides." p. 246. (Cambridge Univ. Press.)
MONTGOMERY, C. G., DICKE, R. H., and PURCELL, E. M. (1948).—"Principles of Microwave Circuits." p. 393. M.I.T. Radiation Laboratory Series, Vol. 8. (McGraw-Hill: New York.)
MUMFORD, W. W. (1949).—*Bell Syst. Tech. J.* **28**: 608.
PRIME, H. A. (1952).—*Aust. J. Sci. Res. A* **5**: 592.
ROSE, D. J., KERR, D. E., BIONDI, M. A., EVERHART, E., and BROWN, S. C. (1949).—M.I.T. Radiation Laboratory Tech. Rep. No. 140.

THE PROPERTIES OF LIQUID PHASE EMBRYOS AT THE INTERSECTIONS OF PLANE SOLID SURFACES

By R. G. WYLIE*

[*Manuscript received July 28, 1952*]

Summary

In order to calculate the probabilities of nucleation of liquid droplets at different types of site on a solid surface, the properties of embryonic droplets which may exist in complete thermodynamic equilibrium at those sites must be known. The general properties of liquid embryos formed on a plane solid surface, or at lines or points of intersection of plane solid surfaces, are considered. It is shown that, although an edge free energy associated with the boundary line may substantially affect the properties of embryos at small contact angles, the effect is probably not large, for embryos of the sizes of interest, when the contact angle is larger than about $\pi/4$.

The areas, volumes, total surface free energies, and free energies of formation are found for embryos at these sites as functions of the contact angle, any edge free energies being neglected. The extension to the formation of bubbles at plane solid surfaces in a pure liquid is indicated. The results are applied in a following paper to the kinetics of condensation of a vapour at an imperfect crystalline surface.

I. INTRODUCTION

The formation at phase boundaries of elements of new phases, first considered systematically by Gibbs (1906), has since remained largely a neglected subject, in spite of the fact that most phase changes begin at boundary surfaces rather than in the interior of a homogeneous phase. In the present paper the equilibrium properties of localized liquid phase embryos, formed on a plane solid surface or at the intersections of plane solid surfaces, are considered. The extension to the formation of bubbles at plane solid surfaces in a pure liquid is indicated. The results obtained have been applied to the kinetics of condensation of a vapour at an imperfect crystalline surface (Wylie 1952*b*).

II. GENERAL PROPERTIES

Assuming that the contact angle has a unique value at each plane surface, it is easily seen that liquid embryos which exist in complete equilibrium on a plane, or at a line, or a point of intersection of a set of planes, possess spherical free surfaces. Except when they are extremely small, their shapes are similar to those of macroscopic phases in similar circumstances unless, of course, the latter are distorted by gravitational forces. In fact, if the free energy of the whole system, including the fluid phase with which a localized embryo is in stable or unstable equilibrium, is expressed in terms of the shape parameters and the size, the condition that the free energy should be stationary reduces to

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

statements that the free surface is spherical, that its radius is given by Thomson's equation, and that Neumann's equation holds at each segment of the boundary line.

With the assumption that the embryonic phase is incompressible Thomson's (1888) equation may be written in the form

$$\mu_\beta - \mu_\alpha = \frac{2\gamma}{r}v, \dots\dots\dots (1)$$

where μ_β and μ_α are respectively the thermodynamic potentials of the original and new phases calculated at the external pressure, γ is the surface free energy, v is the molecular volume in the new phase, and r is the radius of curvature of the free surface. Neumann's equation for the i th segment of the boundary line may be written

$$\gamma \cos \theta_i + \gamma_i' = 0, \dots\dots\dots (2)$$

where γ_i' is the excess of the embryo-substrate interfacial free energy over that of the interface between the substrate and the original phase and θ_i is the contact angle.

When the problem is approached in this way, the concept of the internal pressure of the embryo is avoided so that the result is equally applicable to an embryo the rigidity of which precludes internal flow in the time intervals under consideration. Crystalline embryos, which are not entirely isotropic even when cubic in structure, possess surface free energies which depend on the crystallographic direction of the surface and consequently do not form spherical free surfaces when in equilibrium. The calculations outlined below could apply to these, at best, in the sense of being equivalent. In the considerations which follow it will be assumed that the boundary planes are crystallographically equivalent so that only a single contact angle, θ , is involved; this is a simplifying but not a necessary assumption.

As the thermodynamic potential difference, $\mu_\beta - \mu_\alpha$, is increased, the equilibrium radius of an embryo decreases, a stage eventually being reached when, over a significant fraction of one or more of the bounding surfaces, the natural thickness of the surface becomes comparable with the "thickness" of the embryo measured along the normal to the surface. At this stage it is no longer sufficient to describe the superficial free energy simply in terms of uniform surface free energies. In some circumstances an approximate correction could be included in the form of a linear free energy of the boundary line at which the free surface of the embryo intercepts a plane surface. Neumann's equation would then be replaced by

$$\gamma \cos \theta + \gamma' + \frac{\sigma}{a} = 0, \dots\dots\dots (3)$$

where σ is the linear free energy and a is the radius of curvature of the boundary line. The contact angle is then a function of the size. Moreover, σ is a function of θ and may be expected to increase as θ is reduced, at first slowly and then rapidly as θ approaches zero. Equation (1) is not affected provided the radius r relates to a portion of the free surface where the embryo is still thick. The

analogous situation for a crystalline embryo implies the failure of Wulff's (1901) theorem (von Laue 1943 ; Burton, Cabrera, and Frank 1951), which holds for a free crystallite or a crystallite on supporting surfaces only if the interfacial free energy is a function of the local crystallographic direction and not the local curvature. The equilibrium shape is then no longer independent of size. Fortunately, the natural thickness of the surface of a homopolar liquid is probably not much greater than a molecular distance whilst for Kossel's model of a simple-cubic crystal with simply nearest-neighbour interactions it is zero.

When the original phase is a vapour, a rough idea of the magnitude of σ can be obtained for a liquid bound by Lennard-Jones type forces. This may be done in two ways giving lower and upper limits between which σ probably lies. In the first approach and with Frenkel's (1946) conception of the structure of a liquid in mind, the liquid is regarded as possessing a regular simple-cubic lattice and a contact angle of $\pi/2$ is considered. If the simplest assumption is made, namely, that only nearest-neighbour interactions need be considered, the edge potential energy is found to be zero since a molecule in the boundary line simultaneously plays the roles required of it by both the surfaces of which it forms part. A similar situation exists with respect to the vibrational contributions to surface entropy, and σ may be taken to be zero.

However, to be realistic, at least the next-nearest-neighbour interactions must be considered. These act through distances $\sqrt{2}d$, where d is the lattice spacing, and are associated with an energy about 10 per cent. of that of the nearest-neighbour bond. If the deficits of interactions experienced by a molecule in the liquid-vapour surface and a molecule in the liquid-substrate interface, relative to a molecule in the interior of the liquid, are together discounted from the deficit for a molecule in an edge position, a net deficiency of one next-nearest-neighbour interaction is found. This represents double the edge potential energy. The contribution of the entropy to the edge free energy is probably of the same order. For liquids not near their critical temperatures the contribution of entropy to surface free energy is significant but not large in relation to surface energy. Since the energy of a nearest-neighbour interaction, U , may be associated with twice the surface energy per molecule, σ is estimated to be

$$\sigma \sim \frac{1}{2} 10^{-1} U/d \sim 10^{-1} \gamma d.$$

Using equation (3), this result is found to correspond to an increase of 1 in the contact angle, $\pi/2$, when $a \approx 5d$, which is of the same order as the radii of curvature of interest in the theory of condensation on solid surfaces. For water at 0°C. a value for σ of 2×10^{-7} ergs/cm. is obtained. The argument loses its significance if the contact angle is decreased much below $\pi/4$.

In the second approach an attempt may be made to estimate the maximum edge potential energy per molecule which is compatible with an edge molecule remaining in position. Since the surface molecules in a liquid model based on a simple-cubic lattice must be presented with occasional opportunities to evaporate with the rupture, on the average, of three nearest-neighbour bonds, it follows that the energy of removal of an atom in the boundary line cannot be very much less than the heat of vaporization of the liquid (cf. Stranski 1949). The maximum

edge potential energy may therefore be taken to be, say, $\frac{1}{2}U$ per molecule, that is, γd per unit length. This value would result in a considerable effect on the contact angle and work of formation of embryos of relevant size, particularly for small values of the contact angle. For water at 0 °C., σ would be approximately 2×10^{-6} ergs/cm. On account of the way in which this value is derived, it is necessarily of the same order as the edge free energy of an adsorbed monomolecular plate (see Wylie 1952*a*). This argument, however, represents an extreme view. It may be concluded that the effect of the edge free energy is probably significant but not large for contact angles greater than about $\pi/4$ and that for smaller contact angles the magnitude of the effect is unknown, probably becoming very considerable as the contact angle approaches zero. In the following calculations σ is neglected.

Volmer (1939) has pointed out that the work of formation of an embryo of equilibrium size is $\frac{1}{3}F_s$, where F_s is the total superficial free energy. However, the result is not of general validity but holds for embryos which are not too small only if the shape of the embryo is independent of its size, as the following simple derivation shows. With preservation of shape and at constant external pressure, say, the Gibbs free energy of the whole system which includes the embryo is given by a relation of the form

$$G = ar^3 + br^2 + c, \quad \dots\dots\dots (4)$$

where r is a characteristic length of the embryo and a , b , and c are constants, a and b depending on the parameters which define the shape. At equilibrium G is a minimum, that is,

$$3ar^2 + 2br = 0. \quad \dots\dots\dots (5)$$

Substituting (5) in (4) and subtracting c , the work of formation is found to be $\frac{1}{3}br^2$ or $\frac{1}{3}F_s$, which is made up of $-\frac{2}{3}br^2$ arising from the volume free energy and br^2 for the surface. For example, equation (4) could relate to a liquid embryo formed on a plane solid surface but not to one formed on a solid sphere where the shape would vary with size.

It will be apparent that most of what has been said is not restricted to the formation of embryos from a vapour phase. In particular, the results are applicable to bubble formation at solid plane surfaces immersed in a pure liquid, the only modifications necessary being the re-definition of θ to be the supplement of the contact angle and the replacement of equation (1) by

$$kT \log \frac{p_0 + 2\gamma/r}{p} = v(p_0 - p),$$

in which k is Boltzmann's constant, T is the absolute temperature, p_0 is the pressure in the liquid, and p is the saturation vapour pressure of the liquid. Not only is the discussion of the magnitude of σ applicable also in this case but, since the equilibrium vapour pressure in the bubble is independent of r at constant p_0 , the argument of equations (4) and (5) holds and hence all the following considerations are valid.

III. THE PROPERTIES OF EMBRYOS IN PARTICULAR CASES

(i) The areas, volumes, total surface free energies, and free energies of formation will be found as functions of contact angle for spherical embryos on a plane surface, at the angle formed by two intersecting planes and at the corner formed by three intersecting planes. In the three cases the shape is independent of size provided the planes are of sufficient extent. The last two types of site, which may be referred to by the terms "step" and "corner", will be identified

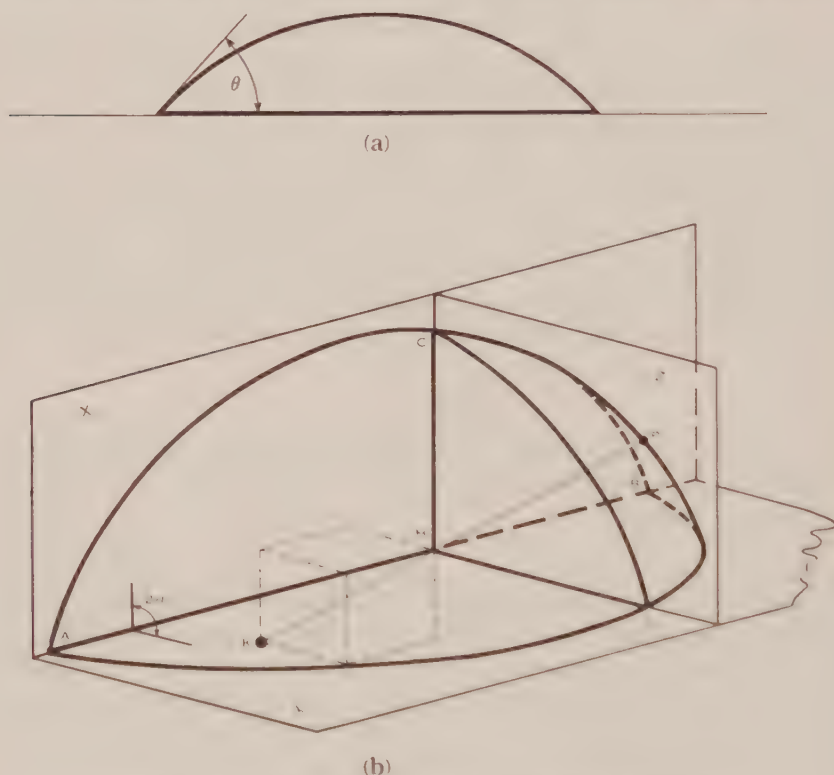


Fig. 1.—The forms of embryos at plane solid surfaces.

(a) Cross section of an embryo on a plane.

(b) Embryos $ACBD$ at the intersection of two planes X and Y and BCD at the intersection of three planes X , Y , and Z (isometric).

in a following paper (Wylie 1952*b*) with two typical features of a crystalline surface which are probably of special importance as nucleation sites.

The three cases which will be considered are represented in Figure 1, the first in cross section and the other two in perspective. The embryo at the line, AB , of intersection of the two planes X and Y is bounded by the plane areas ABC and ABD and by the spherical area $ACBD$. The embryo at the corner, H , of intersection of the three planes X , Y , and Z is bounded by the three plane areas HBC , HCD , and HDB and by the spherical area BCD . The centre of curvature of the spherical areas is K .

The work of formation of an equilibrium embryo on a plane surface has been given by Volmer (1929) as

$$\Delta F = (\pi/3)(2 + \cos \theta)(1 - \cos \theta)^2 r^2 \gamma. \quad \dots\dots\dots (6)$$

The spherical and plane bounding areas and the volume are given respectively by

$$\left. \begin{aligned} A &= 2\pi r^2(1 - \cos \theta), \\ B &= \pi r^2 \sin^2 \theta, \\ V &= (\pi/3)(2 + \cos \theta)(1 - \cos \theta)^2 r^3. \end{aligned} \right\} \quad \dots\dots\dots (7)$$

(ii) To find the corresponding results for embryos at a step and at a corner it is convenient first to calculate the areas of the spherical portions of the surfaces. For an embryo at the intersection of two crystallographically similar planes which cut at an angle 2α this area is found to be

$$\left. \begin{aligned} A_1 &= 4r^2 \int_0^{\arccos(\cos \theta / \sin \alpha)} (\alpha - \beta) \cos \varphi d\varphi, \\ \sin \beta &= \frac{\cos \theta}{\cos \varphi}, \end{aligned} \right\} \quad \dots\dots\dots (8)$$

whilst for an embryo at a corner formed by three similar planes, two of which cut at an angle 2α and are both cut by the third plane orthogonally, the area is

$$B_1 = 2r^2 \int_{(\pi/2) - \theta}^{\arccos(\cos \theta / \sin \alpha)} (\alpha - \beta) \cos \varphi d\varphi, \quad \dots\dots\dots (9)$$

where β has the same meaning as in (8). These results differ only in one of the limits of integration. Consequently, if the integrand is plotted, the two integrals can be found by planimeter integration of the same curve.

The plane surfaces bounding the embryo at a step are each equal to

$$\left. \begin{aligned} A_2 &= r^2 \sin^2 \theta (\psi - \tfrac{1}{2} \sin 2\psi), \\ \sin \psi &= \frac{\sin [\arccos(\cos \theta / \sin \alpha)]}{\sin \theta}. \end{aligned} \right\} \quad \dots\dots\dots (10)$$

At the corner, the two similar plane areas are equal to

$$\left. \begin{aligned} B_2 &= \tfrac{1}{2} r^2 \sin^2 \theta [(\psi - \eta) + \tfrac{1}{2}(\sin 2\psi - \sin 2\eta) - 2 \cos \psi (\sin \psi - \sin \eta)], \\ \sin \eta &= \cot \theta, \end{aligned} \right\} \quad \dots\dots\dots (11)$$

and where ψ has the same meaning as in (10). The plane surface which cuts these areas orthogonally has an area

$$B_3 = r^2 \sin^2 \theta \left[(\alpha - \eta) - \frac{\cot \theta}{\sin \alpha} \sin (\alpha - \eta) \right], \quad \dots\dots\dots (12)$$

where η has the same meaning as in (11). When $\alpha = \pi/4$, (11) and (12) become identical, as symmetry demands.

(iii) The volumes are easily found with the aid of the following device. Taking the case of an embryo at a step, an infinitesimal projection of the embryo is imagined in which the radius vectors of all points, drawn from K (Fig. 1),

are increased in the same ratio infinitesimally larger than unity. The change in volume is

$$dV = (A_1 - 2A_2 \cos \theta) dr.$$

Writing

$$A_1 - 2A_2 \cos \theta = 3gr^2, \quad \dots \dots \dots (13)$$

the result

$$V = gr^3, \quad \dots \dots \dots (14)$$

follows immediately.

Similarly, the volume of the embryo at the corner is given by (14) where

$$3gr^2 = B_1 - 2B_2 \cos \theta - B_3 \cos \theta. \quad \dots \dots \dots (15)$$

The method is, in fact, general for an embryo of constant shape bounded by a system of planes (or, indeed, a developable surface) and a spherical free surface. In general

$$V = \frac{1}{3}(C_1 - C_2 \cos \theta_{12} - C_3 \cos \theta_{13} - \dots)r = gr^3, \quad \dots \dots (16)$$

where C_1 is the spherical area, the remaining C 's are the plane areas, and the angles are the respective contact angles. The reduced volumes, g , of the embryos at steps and at corners are plotted in Figure 2 together with a curve representing the third member of equations (7).

(iv) The free energies of formation, which have been shown to be one-third of the total superficial free energies, are easily obtained. For the embryo at a step

$$F_s = A_1\gamma + 2A_2\gamma',$$

which, on using Neumann's equation, becomes

$$F_s = \gamma(A_1 - 2A_2 \cos \theta),$$

or

$$F_s = 3gr^2\gamma. \quad \dots \dots \dots (17)$$

This result is obviously general, g being defined by (16). The work of formation is thus given by

$$\Delta F = gr^2\gamma. \quad \dots \dots \dots (18)$$

Also,

$$\frac{V}{r} = \frac{\Delta F}{\gamma}. \quad \dots \dots \dots (19)$$

(v) Turnbull (1950) has considered the problem of the retention above the melting point of crystal embryos in conical and cylindrical cavities. The embryos are treated as if they are fluid in character. Turnbull gives an expression for the free energy of formation in a cone of an embryo which is equilibrated as far as Neumann's equation is concerned but not with respect to size. He also gives a complicated criterion that the total superficial free energy should be negative, which can be shown to reduce to the condition that the contact angle should be less than the complement of the semi-angle of the cone. In order to compare the properties of a conical hollow with those of a corner, the volume of the embryo and its free energy of formation will be given.

The volume of the embryo is found to be

$$\begin{aligned} V &= (\pi/3) \{ \cos^3(\alpha + \theta) \cot \alpha - [2 + \sin(\alpha + \theta)][1 - \sin(\alpha + \theta)]^2 \} r^3, \\ \text{or} \quad V &= (\pi/3) \{ \cos^2(\alpha + \theta) \cos \theta \operatorname{cosec} \alpha - 2[1 - \sin(\alpha + \theta)] \} r^3, \end{aligned} \quad (20)$$

where θ is the contact angle and α is the semi-angle of the cone. However, the shape of the embryo does not depend on its size so that equations (14) and (18) are valid. In fact, the conical surface, being developable, may be regarded as made up of an infinite number of plane triangles. The free energy of formation is thus given by (18) where g is defined by (14) and (20).

Separating ΔF into volume and surface terms according to (4) and (5)

$$\Delta F = -2g\gamma r^2 + 3g\gamma r^2. \quad (21)$$

If for a liquid embryo the variation of ΔF with radius at constant thermodynamic potential difference is required it is easily obtained from (1) and (21); thus

$$\Delta F = -g \frac{\mu_\beta - \mu_\alpha}{v} r^3 + 3g\gamma r^2, \quad (22)$$

in which r is positive or negative depending whether the free surface is convex or concave from the outside. If the equilibrium embryo presents a concave free surface then the thermodynamic potential inside the embryo, calculated at the external pressure, exceeds that outside by an amount $(2\gamma/r)v$ and the radius in equation (1) must be regarded as negative. The equilibrium is obviously then stable; the volume term in (4) is positive and the surface term negative. The geometrical criterion that the radius is infinite is simply $\alpha + \theta = \pi/2$, the total superficial free energy being negative if $\theta < (\pi/2) - \alpha$. Mathematically, exactly the same result is reached if the condition is expressed that g is negative. The reduced volume of an embryo in a cone is plotted in Figure 2 for $\alpha = \pi/4$.

(vi) It will be observed that the volume curves for $\alpha = \pi/4$ do not extend to contact angles smaller than $\pi/4$. The implication is that when θ is reduced below the value $\pi/4$ the embryo is not bounded along the line or lines in which planes meet at an angle $2\alpha = \pi/2$. In such a case no localized embryo can exist at the step or corner. When two planes, or two of three planes, meet at an angle 2α , an embryo will spread indefinitely along the line of intersection as soon as θ is reduced to a value less than $(\pi/2) - \alpha$. The fulfilment of this condition is associated with failure of the mechanical equilibrium, which is governed by Neumann's equation, independently of the phase equilibrium governed by equation (1). For an embryo in a cone, on the other hand, the mechanical equilibrium cannot fail. In general, neglecting any linear free energies, σ , when θ is reduced to a value less than $(\pi/2) - \alpha$, a condensate will form along the whole of the corresponding line of intersection before the thermodynamic potential of the mother fluid is raised to the equilibrium value for the macroscopic transformation. On the other hand, the limiting value of θ corresponding to the free surface of a corner embryo being plane is given by

$$\cot \theta = \sin \alpha,$$

which always gives a value for θ larger than $(\pi/2) - \alpha$. When $\alpha = \pi/4$, θ is $54^\circ 48'$.

It has been pointed out by several authors, both in the fields of phase nucleation and adsorption hysteresis, that when an embryo forms in a surface pit or pore, such as the conical hollow discussed by Turnbull, it may not be able

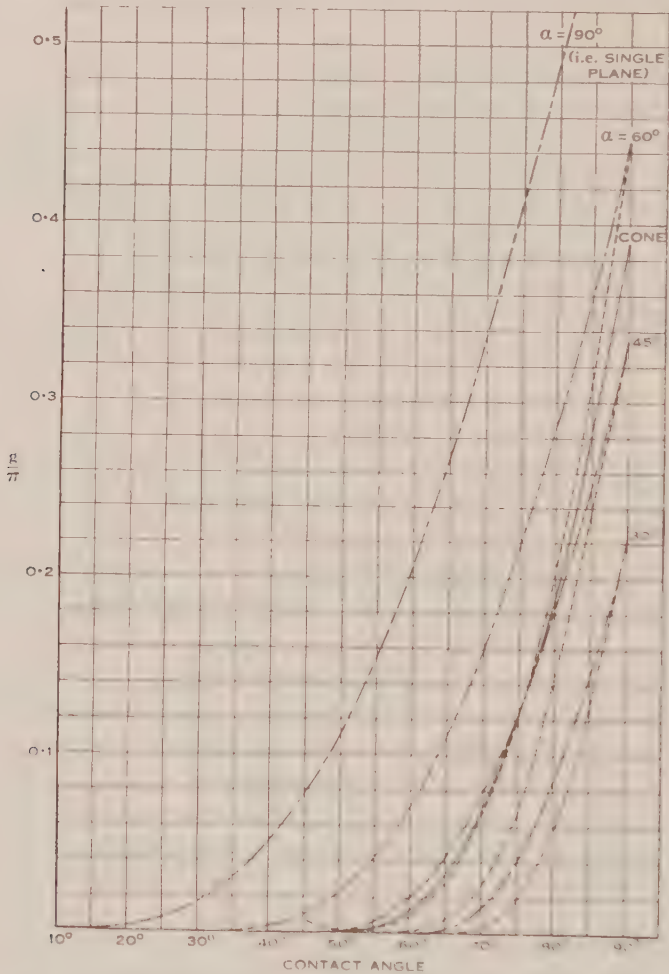


Fig. 2.—The reduced volumes, g , of embryos on plane surfaces and in a cone.

- — — At the line of intersection of two planes which cut at an angle 2α .
- - - - At the point of intersection of three planes two of which intersect at an angle 2α and both cut the third orthogonally. (Volumes $\times 2$.)
- In a cone of semi-angle 45° . (Volumes $\times 2$.)

to grow beyond the orifice without passing through a stage in which its curvature increases markedly (in the direction of convexity as viewed from outside). The consequence is the existence of a range of stable equilibrium and a hysteresis effect. This is true for a conical hollow and intuitively it must also be so for a

pyramidal hollow. Further, it seems very probable that it is so for a corner formed on a surface by the intersection of two steps of limited height. The quantitative treatment of these cases, however, involves the difficult task of dealing with aspheric free surfaces with complicated boundary conditions.

The embryo which forms at a single step of limited height ($\theta > (\pi/2) - \alpha$) must also become aspheric when, through growth, it touches the top of the step. Whether, in this case, the total curvature can ever increase with increasing size is a geometrical problem which awaits solution. In the analogous two-dimensional case (Wylie 1952*a*) such an increase can occur.

IV. ACKNOWLEDGMENT

The author wishes to thank Miss J. Ward for the planimeter integrations and much of the computational work.

V. REFERENCES

- BURTON, W. K., CABRERA, N., and FRANK, F. C. (1951).—*Philos. Trans. A* **243**: 299.
FRENKEL, J. (1946).—“Kinetic Theory of Liquids.” (Oxford Univ. Press.)
GIBBS, J. W. (1906).—See “Commentary on the Scientific Writings of J. Willard Gibbs.” Vol. 2. Ch. 5. (Yale Univ. Press: 1936.)
VON LAUE, M. (1943).—*Z. Kristallogr.* **105**: 124.
STRANSKI, I. N. (1949).—*Discuss. Faraday Soc.* **5**: 13.
THOMSON, J. J. (1888).—“Applications of Dynamics to Physics and Chemistry.” (Macmillan & Co.: London.)
TURNBULL, D. (1950).—*J. Chem. Phys.* **18**: 198.
VOLMER, M. (1929).—*Z. Elektrochem.* **35**: 555.
VOLMER, M. (1939).—“Kinetik der Phasenbildung.” (Steinkopf: Leipzig.)
WULFF, G. (1901).—*Z. Kristallogr.* **34**: 449.
WYLIE, R. G. (1952*a*).—*Aust. J. Sci. Res. A* **5**: 288.
WYLIE, R. G. (1952*b*).—*Aust. J. Sci. Res. A* **5**: 628.

THE CONDENSATION OF A VAPOUR AT A CRYSTALLINE SURFACE

By R. G. WYLIE*

[*Manuscript received July 28, 1952*]

Summary

The theory of self-nucleation in a supersaturated vapour has been developed with considerable success by Becker and Döring (1935), but little attention has hitherto been given to condensation at a solid surface.

A theory is given for the nucleation of liquid condensates on a plane surface and at lines and corners formed by the intersection of plane surfaces. It is suggested that the line and corner sites are representative of typical features of real crystalline surfaces. The equilibrium properties and free energies of formation of embryos at these sites, which are required for the nucleation theory, have been given in the previous paper.

The theory shows that the supersaturation for which nucleation proceeds at a just observable rate on a plane surface increases rapidly as the contact angle increases from zero. Unless an edge free energy is introduced, no supersaturation is necessary for nucleation at a line or corner of intersection of plane surfaces when the contact angle is less than a threshold value which depends on the angles between the planes. As the contact angle increases above the threshold value the supersaturation required for observable condensation increases considerably.

The effect of contamination of the substrate surface is considered and the suggestions are made that, if a small amount of soluble contamination is present it is likely to become localized at surface steps and corners, sensitizing them as nucleation sites, and that microscopic observation of incipient condensation on solid surfaces may provide a tool for the investigation of surface structure.

I. INTRODUCTION

Whereas the probability of self nucleation of a vapour to form crystals or liquid droplets has been investigated theoretically by Volmer and Weber (1926) and with considerable success by Becker and Döring (1935), the nucleation of condensation at a solid surface has been given only superficial consideration. The subject has been touched upon by Volmer (1939) and discussed by Frenkel (1946) but the considerations of these authors are based on a simple model of a solid surface which is no longer tenable and no quantitative kinetic theory has been given for the problem, which is complicated by the linear boundary and by surface diffusion. These aspects do not arise in the theory of condensation in the bulk of a vapour.

From a different point of view, writers in the field of surface adsorption have sometimes speculated on the development of three-dimensional condensates on surfaces where the condensation is, in the first instance, visualized as forming by the successive adsorption of monolayers (Cassell 1944). The relationship which these speculations bear to observable effects has not been made clear.

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

In the present paper a theory of the condensation of a vapour on a solid surface is developed, particularly for cases in which a non-wetting liquid is formed. The kinetic theory is given both for the plane surface considered by other authors and for a more realistic model of a crystal surface. Some of the purely thermodynamic aspects have already been discussed in the previous paper (Wylie 1952). The suggestion is made that surface defects play an important part in most cases of condensation at clean crystalline surfaces and that observation of incipient condensation by the ordinary techniques of optical microscopy may provide a tool for the investigation of the detailed structure of crystal surfaces. Except in Section VIII where the effects of contamination are considered, it may be assumed that the surfaces in question are clean in the sense that only the solid and the vapour substance are present.

II. THE MECHANISM OF CONDENSATION

(a) Some Aspects of the Contact Angle

In the previous paper the contact angle was regarded as a directly observable quantity related to the interfacial free energies by Neumann's equation. Some further aspects of the contact angle will now be considered.

A solid surface which has been equilibrated with a vapour generally possesses an adsorbed film of vapour substance. In simple cases this may be visualized as consisting of a gaseous monolayer supported either on the bare substrate or on a further adsorbed monolayer which may be strongly bound. The whole adsorbed film modifies the surface free energy of the solid according to Gibbs's adsorption equation and consequently plays a part in determining the contact angle which the condensed vapour substance makes on the surface. The "contact angle" must therefore be defined to be the angle of contact made by a sufficiently large droplet of the particular liquid on the particular surface when the latter is equilibrated with the vapour.

In examples where appreciable hysteresis of the contact angle is observed, which in clean conditions is probably due to the presence of surface irregularities, as suggested by Shuttleworth and Bailey (1948), the definition may be understood to relate to a droplet small compared with the separation of surface defects and located on a flawless region of the surface. A simple consequence of this interpretation of the contact angle is that, when the liquid does not wet the substrate, the adsorbed film cannot be such as to present a surface indistinguishable from the surface of the bulk liquid. It is convenient to regard as part of the substrate any continuous condensed film which may be present on the solid, probably supporting a further adsorbed layer with gas-like properties.

It is not proposed to consider here condensation which occurs film-wise irrespective of how slowly the liquid is deposited, but arguments will now be adduced which strongly support the view that in drop-wise condensation the deposit has a drop structure at all stages of deposition and does not form embryonic droplets by the collapse of a relatively thick but invisible liquid film. The latter possibility could lead to very different conclusions. Theoretically this possibility is improbable; since the liquid does not wet the substrate, the

free energy decrease associated with the mutual annihilation of two liquid surfaces exceeds the decrease which occurs when a liquid surface is united with the substrate surface. In these circumstances it seems inevitable that the thermodynamic potential of the condensate, forming film-wise, must at some stage during growth exceed that of the saturated vapour. That is, a nucleation process is involved. On quite elementary considerations the probability that any appreciable area of the condensate will surmount such a barrier spontaneously is seen to be entirely negligible. Frenkel (1946) has given a simple theory of film-wise condensation which would certainly lead to this view. However, his calculation is based on a much over-simplified model and, in particular, his assumption, that if the liquid does not wet the solid the intermolecular forces between the substrate and condensate can be neglected, is too optimistic, especially when it is noted that a change in the free energy of mutual annihilation of a liquid and a solid surface of only 1 per cent. is sufficient to change the contact angle from zero to approximately 11° .

Phenomenologically, the film theory is most improbable since it would imply that the contact angle for a drop already formed on a surface would depend on the rate of condensation, decreasing with increasing rate. Whilst this may be so for very great rates of condensation it is not ordinarily observed. Again, if the condensing vapour forms an electrically conducting liquid, the implication would be that the surface conductivity would increase greatly for a small increase in vapour pressure from the equilibrium pressure p_0 . It is evident from the behaviour of electronic valves containing mercury vapour that when mercury condenses on glass this is not so. Finally, the existence of droplets on a surface on which an adsorbed film is present has apparently been observed by Baughman and Mosallam (1938*b*). Therefore, it is reasonable to conclude that condensation which is observed to be drop-wise is in fact usually so from the initial stage of nucleation so that molecular aggregates form continuously at suitable sites on the equilibrated surface. The great majority of these embryos fail to survive the fluctuations which occur and evaporate on account of their surface free energies. When the ambient vapour pressure is sufficiently large, some will, by chance fluctuations, attain the size at which they are in unstable equilibrium with the vapour when they will possess approximately a 50 per cent. chance of growing to a visible size.

(b) The Surface Model

Recent theoretical and experimental work on the growth of crystals (Griffin 1950 ; Burton, Cabrera, and Frank 1951 ; Dawson and Vaud 1951 ; Verma 1951) and the theory of dislocations (see Cottrell 1949 ; Frank 1951) permit the use of a model for the surface which is more realistic than the flawless plane surface formerly adopted. It is necessary to disagree at the outset with Volmer's (1939) view that "The surface of a microscopic particle must always be considered as practically plane". The exact state of the surface must depend greatly on its previous history and particularly the method of production. However, the existence of dislocations in the lattice precludes practical crystal surfaces being perfect over any appreciable area.

The theory which follows is to be understood to apply to the best surfaces which can be produced, which are probably those grown at a low supersaturation from the vapour, those prepared by annealing a surface of suitable crystallographic orientation by heating to a temperature well above the threshold at which surface diffusion becomes appreciable but well below the value at which evaporation becomes significant, and those produced by cleavage. From recent work it would be expected that growth from the vapour would produce surfaces with well-defined growth terraces and that the annealing process would give surfaces on which more or less straight steps or polygonal segments join points at which dislocation lines intersect the surface. Surfaces produced by the fracture of crystals with well-defined cleavage planes may resemble those produced by annealing, but multimolecular steps are bound to be present. Surfaces produced by growth or annealing may, of course, present steps more than one molecular distance in height. As is well known, very large steps are observed when some crystals grow at an appreciable rate from solution.

Grain boundaries and mosaic boundaries can, apart from impurities, be regarded as surfaces in which dislocation lines are concentrated so that lines in which such planes intersect a free surface may, on the molecular scale, present a step structure of considerable complexity.

Probably, an important role is played by corners in which steps, perhaps of molecular height, intersect. Such corners will be formed when dislocations move in a crystal. For example, if a screw axis which is perpendicular to the free surface passes across a growth terrace, a corner is formed in each convolution of the terrace. Again, if an edge dislocation emerges on the surface and disappears, a corner is formed where each existing step is cut by the emergent dislocation. Surfaces produced by cleavage probably already possess corners in abundance. Of course, in other circumstances other blemishes such as etch pits may be present, the nucleating abilities of which may greatly exceed those of simpler surface defects. Practical surfaces are commonly formed in other ways than those mentioned above, notably by vacuum evaporation onto glass, by mechanical polishing, and by chemical attack including electrochemical polishing. The probable structures of surfaces produced in these and other ways will not be discussed here but it must be remarked that it is difficult to imagine surfaces more unsuitable for the investigations suggested below than those produced by vacuum evaporation onto glass. Further, it is problematical whether a really amorphous surface can be produced with the aid of glass or organic polymerizing material; such surfaces as can be produced very probably cannot be regarded as homogeneous smooth planes on the molecular scale.

In the following, the solid surface is visualized as consisting of a very large number of small plane areas of low crystallographic indices. In the simplest case, when the lattice is cubic and the surface is macroscopically a (100) surface, these facets may be (100) faces, not all at the same level, but separated by steps of various heights which intersect in corners. Often the two intersecting steps will pass continuously through such a corner. The angle in which they meet is not restricted but is likely to be dictated by the crystal symmetry. The planes

which connect the different levels at steps and corners will be assumed to be crystallographically equivalent to the surface planes. These connecting planes will, for the present, be assumed to be of large extent so that when a liquid embryo touches the top of a step it has already exceeded the equilibrium size.

(c) *Droplet Formation*

The formation of droplets on the equilibrated crystalline surface will follow a course which in many ways is analogous to the formation of droplets in the body of a supersaturated vapour. However, there are important differences. The most obvious of these is that the shape of an embryo of equilibrium size will be determined by the nature of the site at which it forms, three possible sites being a perfect plane surface, a step, and a corner. Other important differences will appear below.

In accordance with the fact that nucleation is a temperature-activated rate process, the nucleation probability per unit time is governed by an equation of the form

$$J = Ce^{-\Delta F/kT}, \quad \dots\dots\dots (1)$$

where ΔF is the work of activation, k is Boltzmann's constant, T is the absolute temperature, and C is a constant the physical dimensions of which depend on whether J is a probability relating to unit area, to unit length of a line, or to a specific point. If the effect of surface diffusion can be neglected, ΔF is the work of formation of the equilibrium embryo. Equation (1) may be adopted tentatively following the ideas of Volmer and Weber (1926). However, the quasi-thermodynamic theory represented by the equation is based on an entirely arbitrary assumption associated with the time factor which is implicit in the constant C .

When the vapour pressure is raised sufficiently above the saturation pressure, p_0 , to cause droplet formation at many sites according to (1), the condensate grows with the repeated coalescence of neighbouring drops which freshly exposes the surface so that new drops form repeatedly at the original sites. Particularly when the predominating sites are specific points they should be observable through this repetition. Tammann and Boehme (1935) have observed this behaviour in the condensation of water on metal surfaces.

In order to make kinetic calculations along the lines of that of Becker and Döring (1935), the properties of embryos in equilibrium at the nucleation sites must be known. In the previous paper these properties were given for a perfect plane, a step (that is, the angle of intersection of two planes), and the corner formed by the intersection of two rectangular steps. The kinetic treatment may now be given.

III. THE KINETICS OF NUCLEATION

It is convenient first to outline a kinetic treatment of condensation on a surface which is closely analogous to the theory of Becker and Döring. This will be done explicitly for the case of a flawless plane surface, at first neglecting surface diffusion.

A system is imagined in which a plane surface is in contact with a super-saturated vapour of pressure p . The pressure and the extent of the area are such that at any instant a very large number of embryos are formed on the surface by thermodynamic fluctuations whilst only a small fraction of the area is occupied. A dynamical equilibrium is established by removing from the system those embryos which reach a predetermined size of j molecules which appreciably exceeds the equilibrium size, the vapour pressure being maintained constant by the addition of the requisite number of vapour molecules.

Let A_v be the total exposed area of all those embryos per unit area which each contain v molecules. The A_v are dimensionless. On writing q_v for the probability per unit time that a molecule will evaporate per unit area of an embryo containing v molecules and a_0 for the probability per unit time that a vapour molecule will strike unit area of any surface, and assuming an accommodation coefficient of unity, a set of equations

$$J = A_{v-1}a_0 - q_v A_v, \quad v=1, 2, \dots, j, \quad \dots \dots \dots (2)$$

is obtained in which J is the number of nucleation events which occur per unit time per unit area. In the first equation of (2), A_0 is taken as unity. By assumption, A_j is zero.

Let the number of molecules in an embryo of equilibrium size be n . Defining $\beta_v = q_v/a_0$ (this is the reciprocal of the corresponding quantity employed by Becker and Döring), and using Thomson's equation (equation (1) of the previous paper) it is easily found that

$$\beta_v = e^{\alpha(1/r_v - 1/r_n)}, \quad \dots \dots \dots (3)$$

where $\alpha = 2\gamma v/kT$, γ is the surface free energy of the free surface of the embryo, v is the molecular volume in the embryo, r_v is the radius of curvature of the free surface of the embryo, and r_n is that of the embryo of equilibrium size.

Multiplying each of equations (2) by a continued product $\beta_1 \cdot \beta_2 \dots \beta_v/q_v$, the equations take the form

$$\left. \begin{aligned} J/a_0 &= A_0 - \beta_1 A_1, \\ &\dots \dots \dots \\ (\prod_{1}^{v-1} \beta) J/a_0 &= (\prod_{1}^{v-1} \beta) A_{v-1} - (\prod_{1}^v \beta) A_v, \\ (\prod_{1}^v \beta) J/a_0 &= (\prod_{1}^v \beta) A_v - (\prod_{1}^{v+1} \beta) A_{v+1}, \\ &\dots \dots \dots \\ (\prod_{1}^{n-2} \beta) J/a_0 &= (\prod_{1}^{n-2} \beta) A_{n-2} - (\prod_{1}^{n-1} \beta) A_{n-1}, \\ (\prod_{1}^{n-1} \beta) J/a_0 &= (\prod_{1}^{n-1} \beta) A_{n-1} - (\prod_{1}^n \beta) A_n, \\ &\dots \dots \dots \\ (\prod_{1}^{j-1} \beta) J/a_0 &= (\prod_{1}^{j-1} \beta) A_{j-1} - 0. \end{aligned} \right\} \dots \dots \dots (4)$$

When these equations are summed the result is

$$J(1 - \beta_1 - \beta_1\beta_2 + \dots + \prod_{1}^v \beta + \dots + \prod_{1}^{j-1} \beta)/a_0 = A_0 - 0. \quad \dots \dots (5)$$

Becker and Döring have stressed the analogy which exists between equations (4) and (5) and those for the passage of an electric current through a number of resistances in series. The current J is driven through the resistance chain, of which $(\prod_{i=1}^n a_0)$ is a typical resistance, by the potential difference A_0 . Another analogy developed by Zeldovich (1942) considers the formal similarity of a differential equation representing (4) and the equation for one-dimensional diffusion. His analogy extends to the time-dependent differential equation.

The problem now consists in the evaluation of the sum of the "resistances" in equation (5). According to (3), this sum is given by

$$\frac{1}{a_0} \sum_{v=0}^{j-1} e^{\alpha(1-r_1+1-r_2+1-r_3+\dots+(1-r_{j-1}-r_j-r_{j+1}+\dots))} \quad (6)$$

Following Becker and Döring, the sum in the exponent is evaluated after replacement by an integral. Then, with the use of equation (19) of Appendix I the summation of the exponentials can be carried out, an integration procedure being used and certain simplifying assumptions being made the physical significance of which is of considerable importance. It is found that the only significant contribution to the resistance sum, of (6), arises in the neighbourhood of the maximum term, $(\prod_{i=1}^n a_0)$, that is, in the neighbourhood of $v=n$. Since the maximum term itself is found to be closely related to the work of formation of the equilibrium embryo, an important consequence is that the calculation is insensitive to the failure of equation (3), that is, Thomson's equation for the small aggregates which occur in the early stages of growth. The resistance sum, equation (6), is found to be

$$\frac{3n}{a_0} \left(\frac{\pi k T}{3 \Delta F} \right)^{\frac{1}{2}} e^{-\Delta F/kT}, \quad (7)$$

where ΔF is the work of formation of the equilibrium embryo from the vapour at pressure p . Finally, the nucleation rate in the steady state is found to be

$$J = \frac{a_0 A_0}{3n} \left(\frac{3 \Delta F}{\pi k T} \right)^{\frac{1}{2}} e^{-\Delta F/kT}. \quad (8)$$

This result is formally the same as that of Becker and Döring but the meanings of A_0 and ΔF are different.

It is easy to see that analogous treatments could be given for nucleation at a step and at a corner. In the former case J is uniform along a line and A_0 is the effective area of unit length of that line, say $2d$, where d is the diameter of a molecule. In the latter case one is dealing with a specific point and A_0 may be taken to be $3d^2$. In each case ΔF is given by the elimination of r between equations (1) and (18) of the previous paper:

$$\frac{\Delta F}{kT} = 4g \left(\frac{\gamma}{kT} \right)^3 \left(\frac{v}{\log s} \right)^2, \quad (9)$$

where g is the reduced volume, V/r^3 , and s is the supersaturation ratio. The quantity a_0 is given by

$$a_0 = s p_0 \left(\frac{N}{2\pi M k T} \right)^{\frac{1}{2}}, \quad \dots \dots \dots (10)$$

where p_0 is the saturation vapour pressure, N is Avogadro's number, and M is the molecular weight.

The number of molecules in the embryo of equilibrium size is given by (1) and (16) of the previous paper:

$$n = 8g \left(\frac{\gamma}{kT \log s} \right)^3 v^2. \quad \dots \dots \dots (11)$$

Using equation (8), the numerical relationship between J and s may be determined for water vapour at 0 °C. For purposes of calculation, v may be

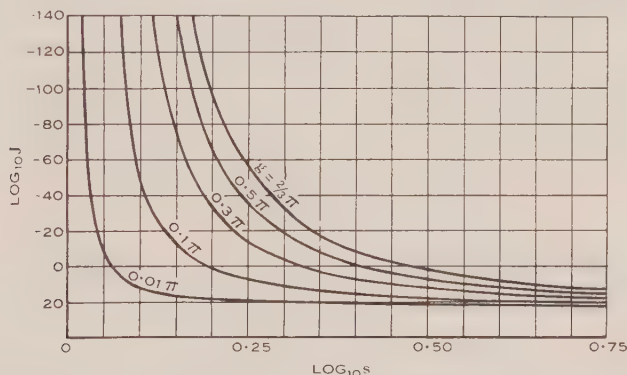


Fig. 1.— Log_{10} of nucleation rate, J per cm^2 per sec., for the condensation of water vapour onto a single plane surface at 0 °C., plotted as a function of log_{10} of supersaturation, s , with the reduced volume, g , as parameter (equation (8)).

replaced by $M/\rho N$, where ρ is the density of the liquid. Figure 1 shows curves of $\text{log}_{10} J$ plotted against $\text{log}_{10} s$ for the three types of condensation site considered. The curves cover a range of g -values, the g -values corresponding to particular geometries having been given in the previous paper.

IV. DISCUSSION OF THE PRELIMINARY CALCULATION

Two important assumptions, which are not necessary for the theory of the self-condensation of a vapour, are implicit in the above argument. The first, which has been discussed in the previous paper, is that in the calculation of ΔF the free energy of the boundary line may be neglected. Any such free energy must increase the contact angle (equation (3) of the previous paper) and hence the volume and work of formation so that nucleation can only be made more difficult. Whilst nothing can be done here to alleviate the need for this assumption which is undoubtedly important for small contact angles and when the equilibrium embryo contains very few molecules, its effect will probably not be serious for the values of n with which we are concerned provided the contact

angle is not much smaller than $\pi/4$. In fact, somewhat unexpectedly, for a given rate of nucleation n increases appreciably as g decreases.

The second important assumption is that only a negligible exchange of molecules occurs through the agency of surface diffusion. With the assumption that the neighbouring surface concentration of adsorbed molecules is unaffected by the embryo for $v > n$, the theory can be re-formulated to include this additional exchange mechanism. Equation (2) becomes

$$J = A_{v-1}[a_0 + 3b_0 d^{-1}(v-1)^{-1/3}] - A_v(q_v + 3q'_v d^{-1}v^{-1/3}), \dots\dots (12)$$

where d is the diameter of a molecule, q'_v is the probability per second per unit length of boundary line that a molecule will leave the embryo by surface diffusion, and b_0 is the probability per second that unit length will be struck by a molecule diffusing on the surface, the associated accommodation coefficient being assumed to be unity. Defining

$$\beta'_v = \frac{q_v + 3q'_v d^{-1}v^{-1/3}}{a_0 + 3b_0 d^{-1}(v-1)^{-1/3}}, \dots\dots\dots (13)$$

and noting that the surface pressure of the gaseous film is proportional to the pressure of the vapour, the principle of detailed balance gives

$$\frac{q'_v v^{-1/3}}{b_0 (v-1)^{-1/3}} \doteq \frac{q_v}{a_0} \doteq \beta'_v \doteq e^{\alpha(1/r_v - 1/r_n)},$$

β'_v also being given by (3). Then

$$\frac{J}{a_0 + 3b_0 d^{-1}(v-1)^{-1/3}} = A_{v-1} - A_v \beta'_v. \dots\dots\dots (14)$$

The argument from this point does not differ from that previously given except in the evaluation of the sum of the partial resistances which now contain factors $(a_0 + 3b_0 d^{-1}(v-1)^{-1/3})^{-1}$ which depend on v . A very good approximation is obtained by treating this factor as a constant with $v = n$. Finally, putting a instead of $n-1$ and neglecting $\pi/3$ compared with the other factors,

$$J = A_1 \left(\frac{a_0 + 3b_0 d^{-1}n^{-1/3}}{3n} \right) \left(\frac{\Delta F^\ddagger}{kT} \right)^3 e^{-\Delta F/kT}. \dots\dots\dots (15)$$

An aspect of a very fundamental nature remains to be considered. In common with the calculation of Becker and Döring the above argument takes no explicit account of the entropy of Brownian motion or communal entropy of the fluctuating embryos. On this account Volmer (1939) included in the expression for the rate of self-nucleation of a vapour a factor of the order of 10^6 which depended on the heat of vaporization. Eucken (1944) has pointed out that this represents a confusion of heat with free energy in connection with the establishment of the first molecule of the aggregate. In the present example the omission becomes evident if an attempt is made to reconcile quantitatively the theory based on the surface concentration of adsorbed molecules (that is, neglecting the first of equations (2)) with the theory based simply on unit area, that is, as first given. In the theory of self-nucleation the omission is closely related to the modification of Thomson's equation put forward by Kuhrt (1951).

who finds by a non-statistical argument that a droplet suspended in a vapour is in equilibrium with a supersaturation given by

$$kT \log s = \frac{2\gamma v}{r} - 4 \frac{kT}{n} \quad \dots\dots\dots (16)$$

The β -values are then not given by (3).

In condensation on a surface, the embryo does not possess the six degrees of freedom of translation and rotation necessary for the validity of (16). When the aggregates are larger than a few molecules they must be completely localized. Nevertheless a communal entropy still exists. The complete partition function for the species of size v is readily obtained if it is agreed that the number of possible sites accessible to an aggregate on an area A is A/d_0^2 , where d_0 is the intermolecular distance in the substrate; certainly within a factor of 2, the stable positions of an aggregate on the surface must be periodic with period d_0 in both dimensions. Since aggregates of different size are in principle distinguishable, the aggregates of one size can be considered as a distinct phase. Regarding the reference state as defined by separated molecules at rest, the partition function for the phase of v -aggregates is

$$Z_v = \frac{(A/d_0^2)!}{N_v!(A/d_0^2 - N_v)!} e^{-N_v \Delta\phi/kT}, \quad \dots\dots\dots (17)$$

where N_v is the number of aggregates of size v and $\Delta\phi$ is the work of formation of a *localized* v -aggregate. Deducing the thermodynamic potential and equating it to that of the vapour, the result is obtained that

$$N_v = (A/d_0^2) \left(\frac{z}{N_0} \right)^{-v} e^{-\Delta\phi/kT} = (A/d_0^2) e^{-\Delta F/kT}, \quad \dots\dots (18)$$

where N_0 and z are respectively the concentration of vapour molecules and the partition function for a single vapour molecule and ΔF is the work of formation of a *localized* v -aggregate from the vapour.

If the aggregates are freely mobile, as could be the case for single adsorbed molecules and perhaps the smallest groupings, the analogous considerations depend on the degrees of freedom assumed to be effective. In the most extreme case, three vibrational degrees of freedom of the stationary aggregate are replaced by two dimensions of gas-like translation on the surface and a free rotation about the normal axis. The partition function of an aggregate becomes

$$z_v = \left(\frac{2\pi m v kT}{h^2} \right) A \left[\frac{4\pi^2 (3/4\pi)^{2/3} m v^{5/3} kT}{h^2} \right]^{1/2} v^{1/3} e^{-\Delta\phi/kT}, \quad \dots\dots (19)$$

where, in the calculation of $\Delta\phi$, there is little choice but to disregard the loss of three internal degrees of freedom. Equation (19) leads to

$$N_v = A \left(\frac{2\pi m kT}{h^2} \right) \left[\frac{4\pi^2 (3/4\pi)^{2/3} m kT}{h^2} \right]^{1/2} v^{1/3} v^{11/6} e^{-\Delta F/kT}, \quad \dots\dots (20)$$

where ΔF is the work of formation of a *localized* aggregate from the vapour.

The condition $\partial N_v / \partial v = 0$ gives a modification of Thomson's equation which is appropriate to this case:

$$kT \log s = \frac{2\gamma v}{r} - \frac{11}{6} \frac{kT}{v}. \quad (21)$$

The essential difference between (20) and (18) is in the absence of the number v from the pre-exponential factor in the latter. The consequence of this is that, when (17) is applicable, Thomson's equation is valid without modification.

If embryos form along a line discontinuity, the arguments consequent on (17) remain valid when $A d_0^2$ is replaced by $L d_0$, where L is the length of the line, Thomson's equation being unmodified. If there is free translation along the line, the result analogous to (21) contains the term $-\frac{1}{2}kT/v$. Embryos located at specific points cannot possess external degrees of freedom and Thomson's equation must be applicable in its usual form. The probability that a specific site is occupied by a v -aggregate is, of course, given by

$$P = e^{-\Delta F/kT}. \quad (22)$$

The significance of these considerations, and some aspects of the nucleation process which are not clearly apparent from the treatment of Becker and Döring, can be appreciated if the theory is re-formulated. The following re-formulation, which draws on certain reliable conclusions of the Becker-Döring type calculation, is of wide application to nucleation problems.

V. A MODIFIED KINETIC THEORY

Consider the argument which extends from equation (2) to equation (8).

Denoting the partial resistance $(\prod_{i=0}^v \beta_i) a_0$ by R_v , equation (5) may be written

$$J = A_0 / \sum_{v=0}^{\infty} R_v. \quad (23)$$

As Becker and Döring have emphasized, practically the whole of the sum $\sum R_v$ arises in the neighbourhood of the term R_n . Following their mathematical procedure the value is found to be

$$\sum R_v = \frac{3n}{a_0} \left(\frac{\pi kT}{3\Delta F} \right)^{\frac{1}{2}} e^{\Delta F/kT} = 3n \left(\frac{\pi kT}{3\Delta F} \right)^{\frac{1}{2}} R_n. \quad (24)$$

In circumstances of interest the ratio $\sum R_v / R_n$ is of the order 10^4 to 10^5 which is barely significant in the present context, especially if attention is focused on the supersaturation, s . In other words, the approximation $\sum_{v=0}^{j-1} R_v \ll R_n$ is quite a

good one whilst, on account of the pronounced maximum which R_v possesses for $v = n$, assuming n to be of the order of 100, the approximation $\sum_{v=0}^{j-1} R_v \approx \sum_{v=-20}^{n+20} R_v$ differs negligibly from (7). These approximations are equivalent to discarding the earlier terms in the summation of equation (5) and hence to replacing the earlier equations of (4) by

$$0 = \left(\prod_{i=1}^v \beta_i \right) A_v - \left(\prod_{i=1}^{v+1} \beta_i \right) A_{v+1}. \quad (25)$$

The physical interpretation is clearly that the aggregates of sizes almost up to $\nu=n$ are practically in statistical equilibrium, equations (25) representing the principle of detailed balance.

Suppose that the fluctuating embryos are in statistical equilibrium up to the size $(n-1)$ and that the embryos which reach the equilibrium size, n , are removed from the system ($j=n$). Equations (25) are then applicable as far as $\nu=n-2$ and the final equation is

$$\left(\prod_1^{n-1} \beta\right) J/a_0 = \left(\prod_1^{n-1} \beta\right) A_{n-1} - 0. \dots\dots\dots (26)$$

These equations are mathematically equivalent to putting R_n for ΣR_ν . The further physical implication is seen to be involved that embryos which reach the critical size necessarily grow to macroscopic droplets, an assumption which represents a factor of 2 out of the factor $\Sigma R_\nu/R_n$.

The theory first given implicitly made use of equations (25) up to, say, $\nu=n-20$, the only function of the equations being to give the equilibrium number of $(n-20)$ -aggregates in terms of the concentration of single molecules. In this role, the equations were used for aggregates down to single molecules which were treated, in effect, as liquid spheres each with a certain volume free energy and with the surface free energy of the bulk liquid and which were localized. This is easily avoided.

From the initial argument, which follows Becker and Döring, the conclusions may be extracted that the deviations from statistical equilibrium do not extend to aggregates much smaller than $\nu=n$ and that the ratio $\Sigma R_\nu/R_n$ is given approximately by (24). The argument may then be reconstructed as follows.

If the population of $(n-1)$ -aggregates is in statistical equilibrium with the vapour, A_{n-1} is given by $\pi d^2(n-1)^{2/3} N_{n-1}/A$, where N_{n-1}/A is given in turn by the appropriate absolute distribution function. For condensation on a flawless plane this is (18) in which the pre-exponential factor $1/d_0^2$ is independent of the vapour pressure and approximately equal to 10^{15} per cm^2 . Then, neglecting surface diffusion and cancelling π with 3, the nucleation rate is given by

$$J = a_0 [d^2(n-1)^{2/3} / (1/d_0^2)] e^{-F/kT} = a_0 (n-1)^{2/3} e^{-F/kT}, \dots\dots (27)$$

where d and d_0 , which differ by only a small factor, have been cancelled.

However, ΔF and hence N_ν are stationary with respect to ν at $\nu=n$ so that, in equation (27), $(n-1)$ may be replaced by n . Also, to take account of the facts that the deviation from statistical equilibrium extends to smaller values of ν than $(n-1)$, and that aggregates which grow to a size slightly larger than the equilibrium size may still evaporate, the ratio $\Sigma R_\nu/R_n$ given by (24) may be incorporated. Finally, since a diffusion field is necessarily associated with a reduced entropy and is hence, in equilibrium, of improbable existence, the reasonable assumption may be made that surface diffusion does not appreciably affect the value of ν below which the aggregates are essentially in statistical equilibrium so that the effect of diffusion is to add a term onto a_0 (that is, in equation (26); see also equation (15)). The result becomes

$$J = \frac{1}{3} (a_0 + 3b_0 d^{-1} n^{-1/3}) n^{-1/3} (\Delta F/kT)^{1/2} e^{-F/kT}, \dots\dots (28)$$

Apart from the surface diffusion term, this expression for J differs from that of equation (8) by a factor of only $n^{2.3}$, which is of the order of 10 to 10^2 , and from that of equation (15) by an indeterminate factor since A_1 remains unknown. The ratio N_{v+1}/N_v obtained from (25) with the use of (3) is not in complete agreement with that given by (18). This is attributable to the treatment of v as a continuous variable in the derivation of (3) and the calculation of ΔF . The β -values are properly given by (18) and (25). Of course, if a result such as (20) were applicable instead of (18) a further source of disagreement would exist.

VI. DISCUSSION OF THE MODIFIED ARGUMENT

It is clear that the modified argument is entirely indifferent to the behaviour of the smaller aggregates both as regards their thermodynamic functions and kinetic properties, for example, accommodation coefficients for the impact of single molecules. In particular, Volmer's (1939) correction for the accommodation coefficients of the smallest aggregates is seen to be irrelevant. The assumption implicit in the theory, that the accommodation coefficient for the aggregate of equilibrium size is unity, is very reasonable: the coefficient for bulk water is believed to be unity. Although some reflection of the impact wave is to be expected at the liquid-solid interface, with due allowance for scattering and for the magnitude of the droplet (n is about 100), the accommodation coefficient cannot differ greatly from that of the bulk liquid and, in any case, J is only proportional to it.

It is very obvious that for large values of J the maintenance of the vapour pressure is, in practice, impossible and growing embryos considerably deplete the vapour concentration in their neighbourhood. However, if a mechanism for the automatic removal of large drops exists at all, it is only necessary to reduce J sufficiently to render the effect negligible. This may entail a considerable reduction of J but only a small reduction in s . The question of the removal of large drops is much more important in systems not containing an inert gas but only the pure vapour. Whilst such systems cannot be used for the verification of the Becker Döring theory of self nucleation in a vapour, they could easily be employed in the investigation of surface condensation. In them, the time required for an embryo to grow to visible size once it has passed the equilibrium size will generally be a very small fraction of a second, especially where the thermal conductivity and diffusivity of the solid substrate and the liquid are relatively large.

Assuming that ΔF is calculated accurately, the greatest uncertainty in the numerical application of equation (28) arises in assessing the magnitude of the surface diffusion term. This almost certainly varies greatly from one case to another. However, it is shown in Appendix I that when the vapour pressure is about 1 mm. and the molecular weight is about 20, the ratio of the term to a_0 is not greater than 10^6 and that even when the vapour pressure is as low as 10^{-3} mm. the ratio does not exceed 10^9 ; for water at 0°C. the appropriate limit is 10^5 . A factor of this magnitude changes the calculated value of s significantly but by no means so much as to vitiate the calculation in which surface diffusion

is completely neglected. For example, it is found that, if $J=1$ per cm.² per sec. corresponds to $s=3$, the factor 10^5 reduces s to approximately 2.7. Again, if the same value of J corresponds to $s=1.50$, then s is reduced to approximately 1.45. Very probably the correct factor is often much smaller.

The modified approach which has been developed is of wide application. It is necessary only to assume that appreciable deviations from statistical equilibrium do not extend to ν -values much smaller than n and to be able to calculate the absolute distribution function which is appropriate. Since the existence of a considerable diffusion resistance or heat-transfer resistance to growth does not obviously invalidate the basic assumption, it would appear that the supersaturation necessary for an observable rate of crystallization from a solution, for example, could be calculated in this way. The effect of the entropy of Brownian motion on the original calculation of Becker and Döring is easily obtained by a similar treatment. The pre-exponential factor of the distribution function is found to contain a factor ν^4 .

VII. THE DURATION OF TRANSIENT EFFECTS

If one of the parameters governing the nucleation rate, for example, the vapour pressure, is changed suddenly, a steady state represented by a set of equations of the form (2) is established only after an initial transient period during which J is a function of ν . The most suitable form of nucleation theory for the consideration of a time-dependent nucleation rate is that of Zeldovich (1942).

Regarding the actual aggregate concentration, N' , and the rate J defined by (2) as continuous functions of both ν and the time t , it is easily found that

$$\frac{\partial N'}{\partial t} = -\frac{\partial N'}{\partial \nu} \dots \dots \dots (29)$$

Equations (2) may be replaced by

$$J = -D \left(\frac{\partial N'}{\partial \nu} - N' \frac{\partial \log N}{\partial \nu} \right), \dots \dots \dots (30)$$

where N is the equilibrium distribution that would obtain if nucleation did not occur, and where the "diffusion coefficient", D , is a function of ν . When the supply of molecules by surface diffusion is neglected, this coefficient is equal to $A'a_0$, where A' is the exposed area of an aggregate, and when only the supply of molecules by surface diffusion need be considered, the coefficient is equal to $L'b_0$, where L' is the length of the boundary line.

Equations (29) and (30) lead to the "diffusion" equation

$$\frac{\partial N'}{\partial t} = \frac{\partial}{\partial \nu} \left(D \frac{\partial N'}{\partial \nu} \right) - \frac{\partial}{\partial \nu} \left(D N' \frac{\partial \log N}{\partial \nu} \right). \dots \dots \dots (31)$$

This time-dependent equation has been considered in connection with nucleation in the body of a vapour by Reiss (1950), Kantrowitz (1951), and Probstein (1951). Although the subsequent omission by these authors of the factor ν^4 , which the Brownian motion contributes to N , represents an error in their case.

in the case of condensation on a surface the factor is not appropriate. By means of simplifying assumptions Probstein reduces (31) to

$$\frac{\partial N'}{\partial t} = \frac{D\partial^2 N'}{\partial y^2} - BN', \dots\dots\dots (32)$$

where B is a positive constant the value of which is easily obtained. Even this equation, however, is very difficult to deal with. Kantrowitz finds that computed times are reduced only by a factor of the order of 2 if the last term in (32) is discarded to give the much simpler conventional diffusion equation. For present purposes the latter equation may be associated with Einstein's well-known formula for the mean diffusion of a unit in time t :

$$y^2 = 2Dt. \dots\dots\dots (33)$$

It may also be noted that the mean first passage time for a diffusing unit (in the present case, a fluctuating embryo) is given by a similar result (Klein 1952).

Thus, to an order of magnitude, a transient which affects the distribution function for all y occupies a time interval $n^2/2D_n$ which for water at 0°C., putting $n \sim 100$ and neglecting surface diffusion, is about 10^{-4} sec. In the same case, a transient which affects the distribution function only for y greater than $a = 10^{-4}$, say, occupies a time $10^2/2D_n \sim 10^{-6}$ sec. It is obvious from these values that in experiments on solid surfaces the steady rate may be regarded as established instantaneously.

VIII. CONCLUSIONS

Values of $\log_{10} J$ for different values of g and s are given for water vapour at 0°C. in Table 1, equation (28) having been used and surface diffusion having been neglected. With the aid of the curves of g plotted against contact angle which have been given in the previous paper, the nucleation probabilities for different sites can be obtained. It will be apparent that the different cases lead to the same formula for J if this quantity is always expressed as a probability per surface site, that is, per area d^2 . The common equation is simply (28) multiplied by d^2 . It will be seen that nucleation on a perfect plane will occur only for a supersaturation s which increases rapidly as the contact angle increases from zero. Standardizing on the value unity for J per cm.^2 per sec., the values of s corresponding to contact angles 10, 30, and 50° on a perfect plane are 1.01, 1.2, and 1.6. With the presence of high steps a nucleation barrier exists only for contact angles greater than $(\pi/2) - \alpha$, where the angle between the planes is 2α , at least in the absence of an edge free energy. For a given contact angle, nucleation is far easier at corners than at steps of the same α value. To determine whether nucleation at specific points will be distinguishable in view of the condensation which may occur on the unbroken plane areas, the number of specific points per unit area must be known as well as the type of defect which they represent. When nucleation is observed reproducibly at a point, if the observed point represents a single site (that is, area $\sim d^2$) the nucleation probability per second is given by (28) multiplied by 10^{-15} . Condensation which is uniformly probable along a line will be evident in the microscope by the alignment of the droplets produced.

If the surface possesses only a minimum of imperfections, such as steps one molecular distance in height, the behaviour can only be estimated. For a single step of monomolecular height, the value of J per site will obviously be intermediate between those for a plane and for a high step and probably a good deal closer to the former. Further, the intersection of two monomolecular steps will be a more favoured site than one on a single step.

TABLE 1

VALUES OF THE NUCLEATION RATE AND OTHER QUANTITIES OF INTEREST IN THE CONDENSATION OF WATER ONTO A PLANE SURFACE AT 0 °C.

The quantities tabulated for various values of the reduced volume, g , and supersaturation, s , are the number of molecules in the embryo of equilibrium size, n , the free energy of formation of the equilibrium embryo, ΔF , tabulated in the form $\Delta F/kT$, the latter quantity multiplied by $\log_{10}e$, and the logarithm to the base 10 of the nucleation rate, J per cm.² per sec.

g/π	s	n	$\Delta F/kT$	$\Delta F/kT \log_{10}e$	$\log_{10}J$
0.01	1.05	14,700	359.0	155.9	-135.6
"	1.10	1,970	94.1	40.9	-20.5
"	1.15	626	43.7	19.0	+ 1.4
"	1.25	154	17.2	7.5	+ 12.9
0.10	1.25	1,540	171.6	74.5	-54.0
"	1.50	256	52.0	22.6	-1.9
"	2.00	51	17.8	7.7	+ 13.0
"	2.50	22	10.2	4.4	+ 16.4
0.30	1.5	769	156.0	67.7	-47.0
"	2.0	154	53.4	23.2	-2.3
"	2.5	67	30.5	13.3	+ 7.7
"	3.0	39	21.2	9.2	+ 11.8
0.50	1.5	1,280	259.9	112.9	-92.1
"	2.0	257	88.9	38.6	-17.8
"	2.5	111	51.0	22.1	+ 0.2
"	3.0	64	35.4	15.4	+ 5.7
2/3	2.0	342	118.6	51.5	-30.6
"	2.5	148	67.9	29.5	-8.5
"	3.0	86	47.2	20.5	+ 0.6
"	5.0	27	22.0	9.6	+ 11.7

It has been shown in the previous paper that, for values of n of the order of a hundred and provided the contact angle is not much less than $\pi/4$, the edge free energy is unlikely to contribute appreciably to the free energy of formation. It is instructive to consider a value of s rather smaller than that for which $J=1$ per cm.² per sec. In Table 1, for $g=0.30\pi$, a value for s of 1.5 is associated with an equilibrium size $n=769$ and a value of $\log_{10}J$ of -47.0. At this value of n , however, the effect of any edge free energy will be considerably smaller

which leaves the statement quite intact, that when $g=0.30\pi$, $s=1.5$ corresponds to an entirely negligible rate of nucleation whether diffusion is considered or not. Bangham and Mosallam (1938*a*) have claimed that, on mica, condensation in cracks does not occur at the equilibrium vapour pressure, p_0 , even when the liquid wets the solid. This suggests that a nucleation barrier is provided by an edge free energy. It is for experiment to determine how far this possibility extends in practice.

The mechanism of condensation can be affected by the presence of contamination, especially if the impurity is soluble. That the effect can be very great is clear from the adsorption measurements of Bowden and Throssell (1951). Fortunately, soluble contamination can be removed by continued washing with the condensate itself as actually occurs in some practical applications. Complete removal, however, will not be possible where the contaminant is continuously formed by a chemical process. If a population of droplets on a surface is evaporated slowly, each drop will cling to a surface defect and, when it finally disappears, will do so at an edge or more probably a corner, irrespective of where it was first formed. A concentration of soluble impurities at steps and corners will probably result. Monomolecular steps and corners will also be favoured in this way. Since the localization of soluble matter at any site is strongly favourable to nucleation at that site, it is seen that a sensitization of the defects as nucleation sites is to be expected. Verma (1951) has observed the preferential condensation of water at steps when breathing onto a carborundum crystal. Probably the effect was enhanced by the presence of minute quantities of soluble matter. Insoluble contamination, including small discrete particles, cannot in general be removed in the true sense but the formation of surfaces which are free of such contamination is quite possible in practice. The vapour, of course, must be free of condensation nuclei. This condition and that of freedom from initial contamination was not fulfilled in the experiments of Tammann and Boehme (1935).

The possibility of using the observation of incipient condensation as a tool for the investigation of surface structure is obvious. A good deal might be learnt about a surface by observing the condensation on it, in an uncontaminated state, of a succession of liquids giving a range of contact angles. The technique would be analogous in some ways to the use of a Wilson expansion chamber for the investigation of condensation nuclei but the practical and theoretical difficulties associated with gas expansion would not be encountered.

IX. REFERENCES

- BANGHAM, D. H., and MOSALLAM, S. (1938*a*).—*Proc. Roy. Soc. A* **165**: 552.
 BANGHAM, D. H., and MOSALLAM, S. (1938*b*).—*Proc. Roy. Soc. A* **166**: 558.
 BECKER, R., and DÖRING, W. (1935).—*Ann. Phys. Lpz.* **24**: 719.
 BOWDEN, F. P., and THROSSSELL, W. R. (1951).—*Proc. Roy. Soc. A* **209**: 297.
 BURTON, W. K., CABRERA, N., and FRANK, F. C. (1951).—*Philos. Trans.* **243**: 299.
 CASSELL, H. M. (1944).—*J. Chem. Phys.* **12**: 115.
 COTTRELL, A. H. (1949).—"Progress in Metal Physics." Vol. 1. Sect. 2. (Butterworths: London.)
 DAWSON, I. M., and VAND, V. (1951).—*Nature* **167**: 476.

- EUCKEN, A. (1944).—"Lehrbuch der Chemischen Physik." p. 1389. (Akademische Verlags-gesellschaft: Leipzig.)
- FRANK, F. C. (1951).—*Phil. Mag.* **42**: 809.
- FRENKEL, J. (1946).—"The Kinetic Theory of Liquids." (Oxford Univ. Press.)
- GRIFFIN, L. J. (1950).—*Phil. Mag.* **41**: 196.
- KANTROWITZ, A. (1951).—*J. Chem. Phys.* **19**: 1097.
- KLEIN, G. (1952).—*Proc. Roy. Soc. A* **211**: 431.
- KUHRT, F. (1951).—*Naturwissenschaften* **38**: 281.
- PROBSTEIN, R. F. (1951).—*J. Chem. Phys.* **19**: 619.
- REISS, H. (1950).—*J. Chem. Phys.* **18**: 529.
- SHUTTLEWORTH, R., and BAILEY, G. L. J. (1948).—*Discuss. Faraday Soc.* No. 3: 16.
- TAMMANN, G., and BOEHME, W. (1935).—*Ann. Phys. Lpz.* **22**: 77.
- VERMA, A. R. (1951).—*Phil. Mag.* **42**: 1005.
- VOLMER, M., and WEBER, A. (1926).—*Z. Phys. Chem.* **119**: 277.
- VOLMER, M. (1939).—"Kinetik der Phasenbildung." (Steinkopf: Leipzig.)
- WYLIE, R. G. (1952).—*Aust. J. Sci. Res. A* **5**: 618.
- ZELDOVICH, J. B. (1942).—*J. Exp. Theor. Phys.* **12**: 525.

APPENDIX I

The Magnitude of the Surface Diffusion Effect

The expression for the probability of nucleation at a surface site, J , contains a factor

$$(a_0 + 3b_0 d^{-1} n^{-1/3}),$$

the two terms of which arise respectively from the influx of vapour molecules and of molecules diffusing on the surface. A maximum value of the ratio, ρ , which the second term bears to the first, will be found in two ways for the particular case of a vapour of molecular weight about 20 at a pressure of 1 mm. and for ordinary temperatures.

In the first approach, b_0 is calculated on the assumption that adsorbed molecules are bound at discrete adsorption sites a distance λ apart and separated by energy barriers of magnitude U . The direct exchange of molecules between the adsorbed film and the vapour, which simply tends to maintain the equilibrium surface concentration, does not require explicit consideration. The quantity b_0 is given by

$$b_0 = \lambda \nu' c' e^{-U/kT},$$

where ν' is the vibration frequency of an adsorbed molecule and c' is the surface concentration of adsorbed molecules. The ratio is then given by

$$\rho = \frac{3\lambda \nu' c' d^{-1} n^{-1/3} e^{-U/kT}}{a_0}.$$

The quantities of doubtful magnitude may now be given extreme values so that an upper limit to ρ is found. Thus ν' cannot exceed the vibration frequency in the interior of a condensed substance and may be given the value 10^{13} per sec., c' cannot exceed the concentration for close packing and may be given the value 10^{15} per cm.², and $e^{-U/kT}$ may be put equal to unity. The other factors may be given their known values, $\lambda \div d \div 3 \times 10^{-8}$ cm., $n \sim 100$, and $a_0 \div 10^{21}$ per cm.² per sec. A maximum value of about 10^7 is found for ρ .

In the second approach the value of b_0 is found on the assumption that the adsorbed molecules possess unrestricted mobility on the surface and constitute a two-dimensional perfect gas. Thus

$$b_0 \doteq c' \left(\frac{kTN}{M} \right)^{\frac{1}{2}},$$

where M is the molecular weight and N is Avogadro's number. When c' is given the value 10^{15} per cm.² a limiting value for φ of about 10^6 is found. It may be concluded that, in practice, when the pressure is about 1 mm. and the molecular weight is about 20, the ratio φ will not exceed 10^6 and is probably much smaller in most cases.

For water at 0 °C. a maximum ratio of about 10^5 is appropriate. The vapour pressure would have to be reduced to the low value of 10^{-3} mm. to increase the upper limit to 10^9 . Since φ varies inversely as the square root of molecular weight, no vapour exists at ordinary temperatures for which the maximum φ must be assigned even a factor of 10 larger than the values given for a molecular weight of 20.

THE FREQUENCY DEPENDENCE OF THE DIELECTRIC PROPERTIES OF DIPOLAR SUBSTANCES

By J. J. O'DWYER* and R. A. SACK*

[*Manuscript received May 25, 1952*]

Summary

A critical review of theories of the frequency dependence of the dielectric constant of dipolar substances is given and inconsistencies are pointed out in previous attempts to account for dipolar interaction. A new theory is presented based on Debye's theory of molecular relaxation and using Onsager's method for treating electrostatic interaction. Fluctuations are considered explicitly both in the orientation of an individual dipole and in the local field acting on it; a distinction is also made between the experimentally measurable time constants and the relaxation times governing molecular processes.

The results are presented in the form of a series development in powers of the dipolar concentration. In first approximation the theory leads to a Debye frequency dependence of the dielectric constant; in second approximation a Debye dependence can be made consistent with the molecular mechanism only for a more restricted (two-position) model.

I. INTRODUCTION

Since Debye first suggested that the dielectric properties of many substances could best be explained by assuming that their molecules carried permanent electric moments, a number of theories have been developed to account quantitatively for the behaviour of polar substances in both constant and variable electric fields. Of these the most fundamental is due to Debye himself (1929), who considers the dipolar molecules acted on by an internal (Lorentz) field. It is found that the bulk properties of the substance can be described in exactly the same way whether it is assumed that the dipoles can rotate freely in all directions as in liquids, or whether they are only able to make transitions between two mutually parallel opposite directions as in certain solids. If the microscopic mechanism is such that, in the absence of dipolar interaction, the dipole orientation would approach its equilibrium value with a single relaxation time τ_μ , then the existence of this interaction makes the bulk properties of the substance describable by a single macroscopic relaxation time τ_M . This means that the frequency-dependent dielectric constant is given by

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + i\omega\tau_M}, \dots\dots\dots (1)$$

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

where ϵ_s and ϵ_∞ denote the dielectric constants at very low and very high frequencies respectively and ω is the angular frequency of the applied field. In Debye's model the relaxation times are connected by the relation

$$\tau_M = \tau \mu \frac{\epsilon_s + 2}{\epsilon_\infty + 2}, \dots\dots\dots (2)$$

A frequency-dependent dielectric constant of the form given by equation (1) may also be derived from purely phenomenological considerations on the assumption that only one kind of relaxation mechanism is operative (Maxwell-Wagner effect). If the absorption *v.* frequency curve is wider than that given by equation (1), the result is conveniently expressed by a superposition of various relaxation times.

Qualitative discrepancies between the predictions of Debye's theory and experimental results have led to revisions of the theory of the static dielectric constant, notably by Onsager (1936) and Kirkwood (1939). However, up to the present no consistent theory has been developed for the frequency-dependent dielectric constant on the basis of the models of either Onsager or Kirkwood. In the attempts that have been made to generalize Onsager's theory to include the dynamic case the statistical nature of his treatment has either not been taken into account at all (Cole 1938; Fuoss and Kirkwood 1941; Bolton 1948) or only inadequately so (Fröhlich and Sack 1944). These theories are critically reviewed below. It should be mentioned that none of them leads to a frequency dependence of the dielectric constant of the form of equation (1).

It is the aim of the present paper to show how a Debye formula for the frequency dependence of the dielectric constant can be derived from a kinetic-statistical treatment of Onsager's model. The interaction between the macroscopic field, the orientation of a particular (spherical) polar molecule, and the influence of all the other dipoles (treated as a continuum) are considered in detail. The field is given by a definite time-dependent value, while the other quantities are subject to thermal fluctuations. It is found convenient to express the results as a series expansion in $(\epsilon_s - \epsilon_\infty) / \epsilon_s$; that is, in a series expansion in dipolar concentration. A first order solution is derived from simple considerations, following which a detailed mathematical analysis of the second order solution is given for a more restricted model.

II. REVIEW OF PREVIOUS THEORIES*

In his original theory of the static dielectric constant of isotropic polar materials Debye (1929), following the Clausius-Mosotti approach, considered the dipoles acted on by an internal (Lorentz) field

$$\mathbf{F} = \mathbf{E} + \mathbf{Q}, \dots\dots\dots (3)$$

* *Note added in Proof.*—Since this paper was written the authors have become aware of a theory of dielectric dispersion by Mandel (1951a, 1951b). Mandel overlooks the statistical nature of Onsager's model by stating that the influence of the reaction field always vanishes, and further fails to distinguish between the various relaxation times. Using arguments similar to those of Section IV (c) and Appendix I he derives a formula for $\epsilon(\omega)$ more complicated than (1). In first approximation this is equivalent to our results; conclusions drawn from Mandel's formula for higher values of ϵ_s are unwarranted.

where

$$\mathbf{Q} = 4\pi\mathbf{P}/3. \quad \dots\dots\dots (4)$$

Here \mathbf{E} is the macroscopic field, \mathbf{P} the polarization of the medium, and \mathbf{Q} the field contribution due to \mathbf{P} , which will be referred to as the polarization field. For N molecules per unit volume, each having moment μ_0 and polarizability α , Debye finds

$$\frac{\epsilon_s - 1}{\epsilon_s + 2} = \frac{4\pi N}{3} \left[\alpha + \frac{\mu_0^2}{3kT} \right] = \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} + \frac{4\pi N}{3} \frac{\mu_0^2}{3kT} \quad \dots\dots\dots (5)$$

For an alternating field of angular frequency ω the result is

$$\frac{\epsilon(\omega) - 1}{\epsilon(\omega) + 2} = \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} + \frac{4\pi N}{3} \frac{\mu_0^2}{3kT} \frac{1}{1 + i\omega\tau_\mu}, \quad \dots\dots\dots (6)$$

where τ_μ is the relaxation time governing the microscopic relaxation mechanism. Solving equation (6) for $\epsilon(\omega)$ leads directly to equations (1) and (2). For the model in which the dipole can point in all directions τ_μ is given by

$$\tau_\mu = \zeta/2kT, \quad \dots\dots\dots (7)$$

where ζ is a constant measuring the viscous resistance to molecular rotations, and for the two-position model by

$$\tau_\mu = 1/2W_0, \quad \dots\dots\dots (8)$$

where W_0 is the probability per unit time of a transition from one position to the other in the absence of a field.

Onsager (1936), in his revision of the theory for the static case, considers in detail the behaviour of a single polar molecule in a spherical cavity of radius a within a homogeneous isotropic continuum having the macroscopic properties of the medium. Under the action of an external field \mathbf{E} and a total (permanent and induced) dipole moment \mathbf{m} , the internal field \mathbf{F} is decomposed into a cavity field \mathbf{G} and a reaction field \mathbf{R} , given by

$$\mathbf{F} = \mathbf{G} + \mathbf{R} = \frac{3\epsilon_s}{2\epsilon_s + 1} \mathbf{E} + \frac{2(\epsilon_s - 1)}{2\epsilon_s + 1} \frac{\mathbf{m}}{a^3}. \quad \dots\dots\dots (9)$$

Of these \mathbf{R} is always parallel to \mathbf{m} and hence does not contribute to the orienting couple. Onsager's statistical treatment then gives

$$\frac{(\epsilon_s - \epsilon_\infty)(2\epsilon_s + \epsilon_\infty)}{\epsilon_s} = \frac{4\pi N\mu^2}{kT}, \quad \dots\dots\dots (10)$$

where

$$\mu = \frac{1}{3}(\epsilon_\infty + 2)\mu_0. \quad \dots\dots\dots (11)$$

It is found necessary to assume, in this derivation, that the molecule fills the whole cavity in order that the result may approach the Clausius-Mosotti formula as the permanent dipole moment of the molecules approaches zero.

Criticism has been levelled at Onsager's treatment mainly for the following reasons :

- (i) The "unit cell" model, in which only one molecule is considered in detail and all its surroundings replaced by a macroscopic continuum, represents an unjustifiable simplification. The neglected effects due to the discrete nature, finite size, and short range order of the neighbouring molecules may be more important than the dipolar forces

which are implicitly taken into account in the model. These limitations have been partly removed in a generalization of the theory by Kirkwood (1939), who states his result in such a way that short range interaction can, in principle, be included.

- (ii) The other objection concerns the vanishing of the orienting influence of the reaction field on the central dipole: it has been argued that this vanished only if \mathbf{R} always remained parallel to \mathbf{m} , whereas in fact it cannot instantaneously follow fluctuations of the direction of \mathbf{m} . In the extreme case, for example, the direction of the dipole would vary much more rapidly than \mathbf{R} and the orienting field would be the time average of \mathbf{F} given by (3) and (4), leading to Debye's result (5) (see, for example, Fröhlich 1946). This argument, however, does not take into account that Onsager's treatment involves purely statistical considerations and that for these it is immaterial how the substance approaches its equilibrium. As brought out by Kirkwood's (1939) theory and its systematization by Fröhlich (1948, 1949), equation (9) gives not the unique value of \mathbf{R} for a fixed \mathbf{m} but its statistical average, so that equation (10) is correct within the limitations of the model.

Cole (1938) assumed that Onsager's result could be extended to the case of a periodic applied field by introducing a factor $1/(1 + i\omega\tau_0)$ into that term of the static result which corresponds to polarization due to dipole orientation. This assumption is also made by Bolton (1948) and in a slightly different form by Fuoss and Kirkwood (1941). It appears to be quite an arbitrary assumption and unjustified in that it applies the result of a purely statistical theory to a time-dependent process.

A theory based on kinetic considerations has been developed by Fröhlich and Sack (1944): they take account of the fact that the reaction field does not instantaneously follow the molecular dipole under consideration and obtain a rather complicated expression for the frequency-dependent dielectric constant. For the static case this expression does not reduce to Onsager's formula, which has been shown correct on general statistical considerations. The reason for this discrepancy is that thermal fluctuations of the reaction field and part of the cavity field have not been taken into account, and it is shown below that these will cancel the lagging of the reaction field in so far as application to the static dielectric constant is concerned. A further flaw in the theory is that no distinction has been made between various relaxation times, as done by Debye for his theory (see equation (2)) and as will be done below for an Onsager-type model.

III. THE MODEL

The model to be used in the following calculations is similar to that used by Fröhlich and Sack (1944) but it is discussed below since it is desired to elaborate on certain features of it.

The substance is assumed to consist of a continuous medium which has the real dielectric constant ϵ_∞ independent of the frequency of an external field. In addition, it contains N dipoles per unit volume which are non-polarizable and of magnitude μ given by equation (11). The terms polarization, reaction

field, etc. will therefore only refer to the dipolar contribution. This procedure is entirely equivalent to the more usual one in which each molecule is described by a permanent moment μ_0 and a polarizability α but has the advantage of considerably simplifying the calculations.

The model thus consists of a rigid but orientable dipole of moment μ , at the centre of a sphere of radius a and dielectric constant ϵ_∞ , surrounded by a continuous medium having the assumed macroscopic dielectric properties of the substance. The dipole is assumed to be orientable under the combined action of an applied field and thermal fluctuations in such a way that the statistical average of its (vector) moment vanishes for zero local field. The approach of the average moment towards equilibrium for a constant local field is described by a single relaxation time τ_μ . For the more detailed calculations of Section IV (c) to (e) the nature of the orientation mechanism must be further specified and τ_μ related to this mechanism as in (7) and (8); the case of Debye's two-position model will be considered in detail.

The radius of the spherical cavity is taken to be the volume per polar molecule as given by

$$4\pi a^3/3 = 1/N, \quad \dots\dots\dots (12)$$

which arises, as mentioned above, from the condition that Onsager's equation must approach the Clausius-Mosotti equation as the dipolar concentration approaches zero. However, if the electronic polarizability is accounted for by ϵ_∞ , there would be no grounds for fixing the radius of the cavity on static considerations but, as the following calculations show no reason for change, the cavity size as defined by (12) will be retained.

We shall further assume that the time variation of polarization of the bulk substance is describable by a macroscopic relaxation time constant τ_M . This is equivalent to assuming a Debye-type loss curve for the substance as given by equation (1). It is the purpose of the following calculations to test the consistency of the model and to find relationships between the various relaxation times.

IV. MATHEMATICAL METHOD

(a) *The Local Field*

The essential difference between previous theories and that given in the present paper lies in the treatment of the local field. Following Debye it will be considered as given by equation (3), namely,

$$\mathbf{F} = \mathbf{E} + \mathbf{Q},$$

where \mathbf{E} is the external field and \mathbf{Q} the polarization field, the former having a fixed time dependence and the latter being subject to thermal fluctuations. But in view of Onsager's calculation of the local field (cf. equation (9)) the equilibrium value of the polarization field will be

$$\bar{\mathbf{Q}} = \mathbf{G} + \mathbf{R} - \mathbf{E}, \quad \dots\dots\dots (13)$$

where \mathbf{G} and \mathbf{R} are the cavity and reaction fields respectively, given by

$$\mathbf{G} = \frac{3\epsilon_s}{2\epsilon_s + \epsilon_\infty} \mathbf{E} = (1 + \chi) \mathbf{E}, \quad \dots \dots \dots (14)$$

$$\mathbf{R} = \frac{2(\epsilon_s - \epsilon_\infty)}{\epsilon_\infty(2\epsilon_s + \epsilon_\infty)} \frac{\mu}{a^3} = \psi \frac{\mu}{a^3} \quad \dots \dots \dots (15)$$

If Q has a value differing from its equilibrium value (13) then, as long as \mathbf{E} and μ are kept fixed, it will tend, apart from thermal fluctuations, towards its equilibrium value with its own characteristic relaxation time τ_Q . This relaxation time can be determined by considering a cavity from which the dipole has been removed; then in an alternating external field E we have

$$\frac{Q}{E} = \frac{\epsilon(\omega) - \epsilon_\infty}{2\epsilon(\omega) + \epsilon_\infty} \quad \dots \dots \dots (16)$$

But, since we have assumed a macroscopic relaxation time τ_M for the substance as a whole, equation (1) holds, namely,

$$\epsilon(\omega) - \epsilon_\infty = \frac{\epsilon_s - \epsilon_\infty}{1 - i\omega\tau_M},$$

which on substitution into (16) gives

$$\frac{Q}{E} = \frac{\chi}{1 + i\omega\tau_Q}, \quad \dots \dots \dots (17)$$

where τ_Q is found to be given by

$$\tau_Q = (1 - 2\chi)\tau_M, \quad \dots \dots \dots (18)$$

and χ , which is defined implicitly by (14), is given by

$$\chi = \frac{\epsilon_s - \epsilon_\infty}{2\epsilon_s + \epsilon_\infty} \quad \dots \dots \dots (19)$$

The dimensionless parameter χ refers to the static cavity field in Onsager's sense. It will be convenient to introduce another dimensionless parameter referring to the reaction field. This may be done by defining

$$\Gamma = \frac{\mu R}{kT} = \frac{2(\epsilon_s - \epsilon_\infty)^2}{3\epsilon_\infty \epsilon_s} \quad \dots \dots \dots (20)$$

It is appropriate at this stage to discuss relative magnitudes of terms involved in series expansions in the following calculations. Since χ varies directly with dipolar concentration and Γ with the square of this concentration, the first order calculation will include terms to the first power of χ only, while the second order calculation will include, in addition, terms in χ^2 and Γ . In fact, from Onsager's equation for the static dielectric constant, namely,

$$\epsilon_s - \epsilon_\infty = (1 + \chi) \frac{4\pi\mu^2 N}{3kT}, \quad \dots \dots \dots (21)$$

and from the definitions of χ and Γ we obtain after some manipulation

$$\Gamma = \frac{6\chi^2}{(1 + \chi)(1 - 2\chi)} \quad \dots \dots \dots (22)$$

The interaction of the reaction field and the dipole will affect the results only to the order Γ .

(b) *A Simple Argument for First Order Results*

If only terms to the first power in χ are retained, the reaction field, or more precisely the influence of the dipole on the polarization field, can be neglected, and the cavity field and the vector moment of the dipole can be treated as if they had their average values.

In an alternating external field \mathbf{E} the local field acting on the dipole will be the cavity field only, of which a part \mathbf{E} will be the applied field itself and a part $\mathbf{E}\chi$ will lag the applied field in time in a manner determined by τ_Q , so that

$$\mathbf{F} = \mathbf{E} \left[1 + \frac{\chi}{1 + i\omega\tau_Q} \right]. \quad \dots\dots\dots (23)$$

A Debye-type calculation for the component of polarization in the direction of \mathbf{E} gives

$$\mathbf{P}(\omega) = \frac{1}{1 + i\omega\tau_\mu} \frac{\mu^2 \mathbf{E} N}{3kT} \left[1 + \frac{\chi}{1 + i\omega\tau_Q} \right], \quad \dots\dots\dots (24)$$

and this, in view of our assumptions, must lead to the Debye equation (1) which can be suitably rewritten as

$$\mathbf{P}(\omega) = (1 + \chi) \frac{\mu^2 \mathbf{E} N}{3kT} \frac{1}{1 + i\omega\tau_M}. \quad \dots\dots\dots (25)$$

Using the known relation (18) of τ_M to τ_Q , this immediately gives to the first order in χ

$$\tau_M = (1 + \chi)\tau_\mu. \quad \dots\dots\dots (26)$$

The argument given here only shows that a microscopic relaxation mechanism is compatible with a Debye function (1) of the dielectric constant, provided the relation (2) resulting from Debye's original theory is replaced by (26). It can be shown, however, that, to the approximation considered here, equations (1) and (26) follow of necessity from the model. Since this derivation does not fit the logical sequence of the work as well as that given above it is included as Appendix I.

(c) *The General Static Equations*

For the more detailed calculations which take into account the reaction field the model will be further restricted by assuming that non-dipolar forces tend to orient the dipole into either one of two opposite positions of stable equilibrium only. As far as the dipole itself is concerned, this is the Debye two-position model in which the relaxation time τ_μ is given by equation (8). It can now also be assumed that the direction of the external field is parallel to the positions of stable equilibrium for the dipole under consideration. This avoids the unnecessary complications of vectors in the mathematics, and the final result is given for a polycrystalline material by inserting a factor 1/3 in the result for $P(\omega)$, which is known to arise on averaging dipolar directions relative to a given field direction.

Consider now the case of an applied static field E , for which we define $f(1, Q)$ to be the probability that the dipole considered has the direction 1 and the polarization field the value Q , and $f(2, Q)$ to be the probability that the dipole has the opposite direction 2 and the polarization field the value Q . Then

$$\left. \begin{aligned} f(1, Q) &= \frac{1}{Z} \exp \frac{\mu(E+Q) - \alpha Q^2 + \beta QE}{kT}, \\ f(2, Q) &= \frac{1}{Z} \exp \frac{-\mu(E+Q) - \alpha Q^2 + \beta QE}{kT}, \end{aligned} \right\} \dots\dots (27)$$

where the energy terms in the exponentials of (26) are respectively :

- (i) The energy of the dipole in the total local field.
- (ii) The self energy of the polarization field.
- (iii) The interaction energy between the polarization field and the external field.

The constants α and β are as yet undetermined and depend in general on the size and shape of the cavity and on the macroscopic properties of the surrounding medium. The quantity Z may be regarded as a normalization constant given by

$$Z = \int_{-\infty}^{\infty} 2 \cosh \left[\frac{\mu(E+Q)}{kT} \right] \exp \left[\frac{-\alpha Q^2 + \beta QE}{kT} \right] dQ. \dots (28)$$

For the spherical cavity considered α and β may be determined in the following manner. For a given dipole direction the equilibrium values of Q can be obtained from (27) as

$$\left. \begin{aligned} \bar{Q}(1) &= \frac{\mu}{2\alpha} + \frac{\beta}{2\alpha} E, \\ \bar{Q}(2) &= -\frac{\mu}{2\alpha} + \frac{\beta}{2\alpha} E, \end{aligned} \right\} \dots\dots\dots (29)$$

by putting $\partial f / \partial Q = 0$ in each case. But from equations (13), (14), and (15)

$$\left. \begin{aligned} \bar{Q}(1) &= \psi\mu + \chi E, \\ \bar{Q}(2) &= -\psi\mu + \chi E. \end{aligned} \right\} \dots\dots\dots (30)$$

Then from (29) and (30) we have

$$\left. \begin{aligned} \alpha &= 1/2\psi, \\ \beta &= a^3\epsilon_{\infty}/2. \end{aligned} \right\} \dots\dots\dots (31)$$

Substituting (31) into (28) and performing the integration, we have in view of (15) and (20)

$$Z = 2(2\psi\pi kT)^{1/2} \exp(\Gamma/2) \dots\dots\dots (32)$$

up to and including linear terms in $\mu E/kT$. This is equivalent to ignoring macroscopic saturation effects. In the same way, substituting (31) into (27) and expanding as far as linear terms in the external field, we have

$$\left. \begin{aligned} f(1, Q) &= \frac{1}{Z} \left[1 + \frac{\mu E}{kT} + \frac{a^3\epsilon_{\infty}EQ}{2kT} \right] \exp \left[\frac{\mu Q - Q^2/2\psi}{kT} \right], \\ f(2, Q) &= \frac{1}{Z} \left[1 - \frac{\mu E}{kT} + \frac{a^3\epsilon_{\infty}EQ}{2kT} \right] \exp \left[\frac{-\mu Q - Q^2/2\psi}{kT} \right]. \end{aligned} \right\} \dots (33)$$

(d) *The General Dynamic Equations*

The differential equations for the time-dependent case will be

$$\left. \begin{aligned} \frac{\partial}{\partial t} f(1, Q) &= -W_{12}(Q)f(1, Q) + W_{21}(Q)f(2, Q) - \frac{\partial}{\partial Q} \left[\frac{\partial Q(1)}{\partial t} f(1, Q) - D \frac{\partial f(1, Q)}{\partial Q} \right], \\ \frac{\partial}{\partial t} f(2, Q) &= W_{12}(Q)f(1, Q) - W_{21}(Q)f(2, Q) - \frac{\partial}{\partial Q} \left[\frac{\partial Q(2)}{\partial t} f(2, Q) - D \frac{\partial f(2, Q)}{\partial Q} \right], \\ &\dots\dots\dots \end{aligned} \right\} \quad (34)$$

where $W_{12}(Q)$ and $W_{21}(Q)$ are the transition probabilities per unit time for the $1 \rightarrow 2$ and $2 \rightarrow 1$ transitions respectively in the presence of a polarization field Q . The "diffusion" or "Brownian motion" term in the equations accounts for thermal fluctuations in Q , and the diffusion coefficient D may in general be a function of Q . In the static case the pairs of terms on the right-hand sides of equations (34) must vanish identically so that

$$W_{12}(Q)f(1, Q) = W_{21}(Q)f(2, Q) \quad \dots\dots\dots (35)$$

and

$$\frac{\partial Q(1)}{\partial t} f(1, Q) = D \frac{\partial f(1, Q)}{\partial Q} \quad \dots\dots\dots (36)$$

Substitution of (33) in (35) gives to the first order in E

$$\left. \begin{aligned} W_{12}(Q) &= W_0(1 - \mu E/kT) \exp(-\mu Q/kT), \\ W_{21}(Q) &= W_0(1 + \mu E/kT) \exp(\mu Q/kT), \end{aligned} \right\} \quad \dots\dots (37)$$

where W_0 is the transition probability per unit time in the absence of a field as stated previously. In writing down (37) we have also neglected the possible effect of the polarization field in altering the potential barrier which the dipole must cross between the 1- and 2-positions.

It is found necessary (for reasons which are made clear at a later stage of the calculations) to introduce an anisotropy into the model by considering τ_{QL} to be the relaxation time associated with a polarization field parallel to the dipolar direction and τ_{QT} the corresponding quantity perpendicular to the dipolar direction. It will be assumed that the average time variation of Q is described by τ_Q so that

$$\tau_{QL} + 2\tau_{QT} = 3\tau_Q, \quad \dots\dots\dots (38)$$

since to any longitudinal direction there are two perpendiculars. The introduction of this anisotropy is a purely arbitrary step, but it is not surprising that it is found to be necessary, since Onsager's model applies strictly only to an isotropic material and the introduction of one dipole (to be treated on a microscopic basis) leads to a small anisotropy. This anisotropy is of no consequence in the static case and in the first order time-dependent case, but it cannot be expected that it can still be ignored when the calculations are made more exact.

We assume then that

$$\frac{\partial Q(1)}{\partial t} = \frac{1}{\tau_{QL}} [\bar{Q}(1) - Q] = \frac{1}{\tau_{QL}} [\psi\mu + \chi E - Q]. \quad \dots\dots (39)$$

Also from (33) we have to the first order in E

$$\frac{\partial}{\partial Q} f(1, Q) = \left[\frac{\mu}{kT} - \frac{Q}{\psi kT} + \frac{a^2 \epsilon_\infty E}{2kT} \right] f(1, Q) \quad \dots\dots\dots (40)$$

and, substituting (39) and (40) in (36), we have

$$D = \psi kT / \tau_{QL} \quad \dots\dots\dots (41)$$

for the direction being considered.

(e) *Solution for a Periodic Field*

If the applied field is periodic with time we write

$$E = E_0 \exp(i\omega t) \quad \dots\dots\dots (42)$$

and expand the probability function in power series in E

$$\left. \begin{aligned} f(1, Q) &= g(1, Q) + Eh(1, Q) + \dots\dots, \\ f(2, Q) &= g(2, Q) + Eh(2, Q) + \dots\dots, \end{aligned} \right\} \quad \dots\dots\dots (43)$$

where, by definition,

$$h(1, Q) = [\partial f(1, Q) / \partial E]_{E=0}$$

and

$$h(2, Q) = [\partial f(2, Q) / \partial E]_{E=0}.$$

Substitution of (43) into (34) and collection of terms of appropriate powers of E gives, after a certain amount of algebra,

$$\begin{aligned} i\omega h(1, Q) &= \frac{W_0}{Z} \frac{2\mu}{kT} \exp \left[\frac{Q^2}{2\psi kT} \right] - W_0 h(1, Q) \exp \left(\frac{\mu Q}{kT} \right) + W_0 h(2, Q) \exp \left(\frac{\mu Q}{kT} \right) \\ &\quad - \frac{Z}{\tau_{QL}} \frac{1}{Z} \left(\frac{\mu}{kT} - \frac{Q}{\psi kT} \right) \exp \left(\frac{\mu Q}{kT} - \frac{Q^2}{2\psi kT} \right) - \frac{1}{\tau_{QL}} \left(\frac{\mu}{kT} - \frac{Q}{\psi kT} \right) \frac{\partial h(1, Q)}{\partial Q} + \frac{1}{\tau_{QL}} h(1, Q) \\ &\quad + \frac{kT}{\tau_{QL}} \psi \frac{\partial^2 h(1, Q)}{\partial Q^2}, \quad \dots\dots\dots (44) \end{aligned}$$

$$\begin{aligned} i\omega h(2, Q) &= \frac{W_0}{Z} \frac{2\mu}{kT} \exp \left[\frac{Q^2}{2\psi kT} \right] - W_0 h(2, Q) \exp \left(\frac{\mu Q}{kT} \right) + W_0 h(1, Q) \exp \left(\frac{\mu Q}{kT} \right) \\ &\quad - \frac{Z}{\tau_{QL}} \frac{1}{Z} \left(\frac{\mu}{kT} - \frac{Q}{\psi kT} \right) \exp \left(\frac{\mu Q}{kT} - \frac{Q^2}{2\psi kT} \right) - \frac{1}{\tau_{QL}} \left(\frac{\mu}{kT} - \frac{Q}{\psi kT} \right) \frac{\partial h(2, Q)}{\partial Q} \\ &\quad + \frac{1}{\tau_{QL}} h(2, Q) + \frac{kT}{\tau_{QL}} \psi \frac{\partial^2 h(2, Q)}{\partial Q^2}. \quad \dots\dots\dots (45) \end{aligned}$$

Subtracting (45) from (44) and integrating with respect to Q from $-\infty$ to ∞ gives

$$i\omega P(\omega) = 2W_0 \frac{\mu^2 NE}{kT} \exp \left(-\frac{\Gamma}{2} \right) - 2W_0 I_0 NE, \quad \dots\dots\dots (46)$$

where, by definition,

$$P(\omega) = \int_{-\infty}^{\infty} N \mu E [h(1, Q) - h(2, Q)] dQ, \quad \dots\dots\dots (47)$$

and we have written

$$I = \int_{-\infty}^{\infty} [h(1, Q) \exp(-\mu Q/kT) - h(2, Q) \exp(\mu Q/kT)] dQ. \quad \dots (48)$$

Equation (46) will be the basic equation from which we calculate the polarization. In deriving it we have assumed that $h(1, Q)$ and $h(2, Q)$ and their first and second derivatives with respect to Q , vanish at the limits of integration more rapidly than any given finite power of Q .

The integral of equation (48) can be calculated to any desired order of accuracy by substituting from (44) and (45) for $h(1, Q)$ and $h(2, Q)$ respectively and expanding the exponentials. We set then

$$\begin{aligned} I &= \int_{-\infty}^{\infty} [h(1, Q) - h(2, Q)] dQ \\ &\quad - \int_{-\infty}^{\infty} \frac{\mu Q}{kT} [h(1, Q) + h(2, Q)] dQ \\ &\quad + \int_{-\infty}^{\infty} \frac{\mu^2 Q^2}{2k^2 T^2} [h(1, Q) - h(2, Q)] dQ \\ &= I_0 + I_1 + I_2, \quad \dots (49) \end{aligned}$$

and it remains to evaluate these integrals. By definition,

$$I_0 = P(\omega) / \mu EN, \quad \dots (50)$$

and, after tedious algebra, we find for the other integrals (to the required order in Γ)

$$I_1 = \frac{1}{1 + i\omega\tau_{QL}} \left[\frac{\chi\mu}{kT} \left(1 - \frac{\Gamma}{2} \right) + \frac{\psi\mu}{kT} \frac{P(\omega)}{NE} \right], \quad \dots (51)$$

$$I_2 = \frac{1}{1 + i\omega\tau_{\mu} + 2\tau_{\mu}/\tau_{QL}} \left[\frac{P(\omega)}{NE} \frac{\tau_{\mu}}{\tau_{QL}} \frac{\psi\mu}{kT} + \frac{\psi\mu^3}{2k^2 T^2} (1 - \chi\tau_{\mu}/\tau_{QL}) \right]. \quad \dots (52)$$

Then substituting (50), (51), and (52) into (46) we find for the polarization

$$\begin{aligned} P(\omega) &\left[1 + i\omega\tau_{\mu} - \frac{\Gamma}{1 + i\omega\tau_{QL}} + \frac{\Gamma\tau_{\mu}/\tau_{QL}}{1 + i\omega\tau_{\mu} + 2\tau_{\mu}/\tau_{QL}} \right] \\ &= \frac{\mu^2 NE}{3kT} \left[1 - \frac{\Gamma}{2} + \frac{\chi(1 - \frac{1}{2}\Gamma)}{1 + i\omega\tau_{QL}} - \frac{\frac{1}{2}\Gamma(1 + \chi\tau_{\mu}/\tau_{QL})}{1 + i\omega\tau_{\mu} + 2\tau_{\mu}/\tau_{QL}} \right], \quad \dots (53) \end{aligned}$$

where the factor 3 has been introduced into the denominator of the right-hand side to account for the polycrystallinity of the material (i.e. to average over the angle between the external field and the dipolar direction).

It can easily be verified that, in the static case, equation (53) reduces to Onsager's formula for the static polarization to the required order of magnitude in χ and Γ . This is necessary from general statistical considerations, as pointed out above.

In the frequency-dependent case the various orders of approximation can be obtained from (53) in a straightforward way. Thus, to zero order in χ , we have

$$P(\omega) = \frac{1}{1 + i\omega\tau_\mu} \frac{\mu^2 NE}{3kT},$$

which is the Debye equation obtained when no distinction is made between the internal field and the external field, and it gives $\tau_L = \tau_M$ as is necessary. This result is obtained by taking only the first integral of (49).

To first order in χ equation (51) yields the result already derived by a simple argument in Section IV (b).

To the second order in χ (i.e. to the first order in Γ) it is convenient to assume a trial form

$$T_M = (1 + \chi + x\Gamma)\tau_\mu \quad \text{..... (54)}$$

for the consistency relation, where x is to be determined. Using this trial solution we find that (53) reduces identically to the Debye-Onsager equation (25) to the required order of approximation if $x = \frac{1}{2}$, that is

$$\tau_M = (1 + \chi + \frac{1}{2}\Gamma)\tau_\mu, \quad \text{..... (55)}$$

and in addition

$$\tau_{QL} = (1 - 6\chi)\tau_M, \quad \text{..... (56)}$$

which from (18) and (38) gives

$$\tau_{QL} = \tau_M \quad \text{..... (57)}$$

Equation (55) is the consistency relation which we set out to determine.

V. DISCUSSION

The results derived in Section IV (b) and in Appendix I show that, for substances with spherical molecules and in the absence of strong short range forces, a relaxation mechanism with a single time constant τ_L leads in fact to a simple Debye dependence (1) of the dielectric constant on frequency. These results, which represent a first order approximation in χ and neglect terms in χ^2 and Γ , are therefore directly applicable to pure isotropic substances which are only weakly polar. In this connection it should be noted that for $\epsilon_\infty/\epsilon_s = 5.4$, $\chi = 1.14$ and $\Gamma = 1.30$; while for $\epsilon_\infty/\epsilon_s = 3.2$, $\chi = 1.8$ and $\Gamma = 1.9$, thus effectively stating the range of validity of the first approximation. For dilute solutions of polar molecules in non-polar solvents, the results can be applied without modification if the electronic polarizabilities of the two substances are equal; if they differ, the same conclusions still apply, except that ϵ_∞ , χ , and Γ have to be redefined in terms of the properties of the two constituents.

The calculations of Sections IV (c) to (e) show that the consideration of terms of the order Γ is still compatible with a Debye frequency dependence (1) of $\epsilon(\omega)$, but only if the material surrounding each dipole shows a certain anisotropy in its dynamic properties ($\tau_{QL} \neq \tau_{QT}$). In the two-position model the existence of a preferential axis for any molecule can cause such an anisotropy, but it has not been possible to determine the magnitude of this effect from other

considerations, and hence no check exists for the correctness of the conclusions. Calculations to first order in Γ have also been carried out for the Debye model in which the molecules can rotate freely except for frictional forces, and for a six-position model in which the dipoles can point along three perpendicular axes. In all cases, equation (1) is consistent with the model only on the assumption that variations of \mathbf{Q} parallel and normal to $\underline{\mu}$ are described by different relaxation times τ_{QL} and τ_{QT} , though their difference is smaller than for the two-position model. In each case the value τ_Q in (18) is the harmonic mean of $(\tau_{QL}, \tau_{QT}, \tau_{QT})$. Such an anisotropy is incompatible with the model as it would imply that the material constants of the surroundings depended on the instantaneous direction of the dipole. Hence for these models equation (1) is correct only to first order in χ .

An experimental test for the validity of the theory could be provided by a check of the variation of the macroscopic relaxation time as expressed by (26) and (55) and their modifications for dilute solutions. As long as the polar molecules have identical surroundings, e.g. if they are completely surrounded by non-polar solvent molecules, τ_μ will remain constant whereas τ_M will vary with dipolar concentration. Equation (26) shows, however, that the relative increase in τ_M amounts to only one-third of the relative increase in the static dielectric constant. It is not certain that such a variation (of order 10 per cent.) of the relaxation time would be outside experimental error in view of the strong variation of τ_μ with temperature and the fact that most substances show a certain spread of relaxation times resulting in a dielectric absorption *v.* frequency curve rather broader than a Debye function.

VI. REFERENCES

- BOLTON, H. C. (1948).—*J. Chem. Phys.* **16**: 486.
 COLE, R. H. (1938).—*J. Chem. Phys.* **6**: 385.
 DEBYE, P. (1929).—"Polare Molekeln." (S. Hirzel: Leipzig.)
 FRÖHLICH, H. (1948).—*Trans. Faraday Soc.* **44**: 238.
 FRÖHLICH, H. (1949).—"Theory of Dielectrics." (Oxford Univ. Press.)
 FRÖHLICH, H., and SACK, R. A. (1944).—*Proc. Roy. Soc. A* **182**: 388.
 FUOSS, R. M., and KIRKWOOD, J. G. (1941).—*J. Amer. Chem. Soc.* **63**: 385.
 KIRKWOOD, J. G. (1939).—*J. Chem. Phys.* **7**: 911.
 MANDEL, M. (1951a).—*Physica, 's Grav.* **17**: 799.
 MANDEL, M. (1951b).—*Bull. Socs. Chim. Belg.* **60**: 301.
 ONSAGER, L. (1936).—*J. Amer. Chem. Soc.* **58**: 1487.

APPENDIX I

Direct Derivation of the First Order Solution

The first order results of Section IV (*b*) can be derived directly on the assumption of a microscopic relaxation mechanism without any *a priori* assumptions as to the dielectric behaviour of the bulk substance. We put

$$\varepsilon(\omega) - \varepsilon_\infty = (\varepsilon_s - \varepsilon_\infty)f(\omega), \dots\dots\dots (\text{A1})$$

where $f(\omega)$ is a function to be determined. Debye's calculations show that, in a sinoidal local field $\mathbf{F}(\omega)$, the total polarization is given by

$$\mathbf{P}(\omega) = \frac{\mathbf{F}(\omega)}{1 + i\omega\tau_\mu} \frac{\mu^2 N}{3kT}, \quad \dots\dots\dots (\text{A2})$$

Since the effects of the reaction field can be neglected to the first order in χ , the boundary conditions on the surface of the unit cell yield

$$\frac{\mathbf{F}(\omega)}{\mathbf{E}(\omega)} = \frac{3\varepsilon(\omega)}{2\varepsilon(\omega) + \varepsilon_\infty} = \frac{1 - 2\chi + 3\chi f(\omega)}{1 - 2\chi + 2\chi f(\omega)} = 1 + \chi f(\omega) \quad \dots\dots (\text{A3})$$

to the required order. From (A1) to (A3) and (21) we have

$$(\varepsilon_s - \varepsilon_\infty)f(\omega) = \frac{\varepsilon_s - \varepsilon_\infty}{1 + \chi} \frac{1 + \chi f(\omega)}{1 + i\omega\tau_\mu}, \quad \dots\dots\dots (\text{A4})$$

which, on solving for $f(\omega)$, yields

$$f(\omega) = 1/[1 + i\omega\tau_\mu(1 + \chi)], \quad \dots\dots\dots (\text{A5})$$

which is identical with equations (1) and (26).

TEMPERATURE DEPENDENCE OF THE DIELECTRIC PROPERTIES OF LONG-CHAIN ALIPHATIC ALCOHOLS IN THE SOLID STATE

By J. S. DRYDEN*

[Manuscript received March 11, 1952]

Summary

The dielectric properties of three primary and three secondary long-chain aliphatic alcohols have been investigated within the temperature range of -20 to 70°C . The experimental results are discussed in relation to the theory of Sack on dielectric absorption in linear polar chains and to the conclusions reached in earlier papers on the dielectric properties of these compounds. The apparent activation energies in the primary alcohols are approximately three times those in the secondary alcohols. This indicates either different mechanisms of absorption in the two types of alcohols or, if the mechanisms are the same, significant differences in the energy barriers involved.

I. INTRODUCTION

The dielectric properties, at room temperature, of certain long-chain aliphatic alcohols have been described in three recent publications (Meakins and Mulley 1951; Meakins and Sack 1951; Meakins and Welsh 1951). The authors of these papers have assumed that the model of a linear polar chain capable of reversal, developed theoretically by Sack (1952), is applicable to these compounds.

In order to obtain additional information about the mechanism of absorption, it was decided to investigate the effects of temperature on the dielectric absorption. Measurements were carried out at temperatures in the range -20 to 70°C . on three symmetrical secondary alcohols, 10-nonadecanol ($\text{C}_9\text{H}_{19}\text{CHOHC}_9\text{H}_{19}$), 14-heptacosanol ($\text{C}_{13}\text{H}_{27}\text{CHOHC}_{13}\text{H}_{27}$), and 18-pentatriacontanol ($\text{C}_{17}\text{H}_{35}\text{CHOHC}_{17}\text{H}_{35}$); and on three primary alcohols, *n*-octadecanol ($\text{C}_{18}\text{H}_{37}\text{OH}$), *n*-docosanol ($\text{C}_{22}\text{H}_{45}\text{OH}$), and *n*-hexacosanol ($\text{C}_{26}\text{H}_{53}\text{OH}$).

II. THEORETICAL

Provided certain conditions are fulfilled, Sack (1952) has shown that, if linear dipolar chains of uniform length, consisting of n elementary dipoles, exist in the material, the relative permittivity $\epsilon' - j\epsilon''$ should be described by the following equations:

$$\epsilon' = \epsilon'_{\infty} + \frac{128}{3\pi^3} \cdot \frac{\mu^2 N}{kT} \cdot \frac{(n+1)^2}{1 + \omega^2 \tau_n^2} \quad \dots \dots \dots (1)$$

$$\epsilon'' = \frac{128}{3\pi^3} \cdot \frac{\mu^2 N}{kT} \cdot (n+1)^2 \cdot \frac{\omega \tau_n}{1 + \omega^2 \tau_n^2} \quad \dots \dots \dots (2)$$

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

where μ is the dipole moment of the elementary dipole,

N represents the number of chains per unit volume,

ω is the angular frequency,

k and T , as usual, denote Boltzmann's constant and absolute temperature,

τ_n is the effective relaxation time and is related to τ_0 , the relaxation time of an elementary transition, by the following equation:

$$\tau_n = 2(n+1)^2 \tau_0 / \pi^2. \quad \dots\dots\dots (3)$$

If the dipole chains are not of uniform length, then there will be a distribution of relaxation times rather than a unique value.

The conditions which must be satisfied before these equations are strictly applicable are (i) the polar chains are straight, (ii) the polar chains contain more than about 10 elementary units, (iii) one of the two possible combinations of anti-parallel neighbouring dipoles is excluded, (iv) the permissible states of the whole chain are energetically equal in the absence of a field, and (v) $\epsilon_s - \epsilon_\infty$, the decrease in ϵ in passing through the absorption region, is small compared with ϵ_∞ . In the secondary alcohols condition (v) certainly does not apply and it is not known whether all the others do. However, Sack (1952) considers that deductions from equations (1), (2), and (3) will apply qualitatively to the solid aliphatic alcohols.

An essential difference occurs in the theoretical predictions if there is a small but finite probability of successive dipoles facing each other in the manner excluded by condition (iii). According to Sack (1952) in this case the behaviour for moderately long chains will still be described by equations (1), (2), and (3). But for values of n^2 greater than $e^{J/kT}$ (where J represents the additional energy associated with each pair of neighbours in the less probable combination) the effective relaxation time τ_n will be independent of chain length. A modification of the treatment in an earlier paper by Sack (1946) leads to the conclusion that $\epsilon_s - \epsilon_\infty$ will be proportional to $e^{-J/2kT}$. This treatment also suggests that the relaxation time will be given by the following expression (Sack, personal communication):

$$\tau_n = e^{J/2kT} \tau_0. \quad \dots\dots\dots (4)$$

III. EXPERIMENTAL

The materials used in the investigation were the same as those used in the measurements previously reported (Meakins and Mulley 1951; Meakins and Sack 1951). The method of preparing the specimens has also been reported in earlier publications from this Laboratory (see, for example, Meakins and Sack 1951). In the frequency range 0.5 c/s.-150 kc/s., bridge methods of measurement were used and, between frequencies of 90 kc/s. and 10 Mc/s., a series resonant circuit. The results reported in this paper were obtained using brass electrodes but some measurements in the frequency region covered by bridge methods were also carried out using mercury electrodes. Similar qualitative results were obtained, the values of both ϵ' and ϵ'' being approximately 10 per cent. greater with mercury electrodes. Temperature was controlled in some of the experiments by means of an oil-bath surrounding the electrode and in others by an air stream at the desired temperature passing through the electrode system.

IV. RESULTS

(a) Dependence of Dielectric Properties on Thermal History

Previous measurements have shown that the dielectric absorption in solid secondary alcohols at room temperature is influenced by the ambient temperature during solidification (Meakins and Sack 1951). To obtain further information about the effect of thermal history on the dielectric properties of the solid two experiments were performed on 10-nonadecanol. In the first of these the rate of cooling to room temperature after solidification was varied. The results indicated that the rate of cooling did not have a significant effect on the magnitude of the dielectric absorption at room temperature. However, measurements carried out in the second experiment indicated that this may not have been so if the compound, after solidification, had been cooled very slowly through the temperature range of a few degrees below the freezing point. In this experiment

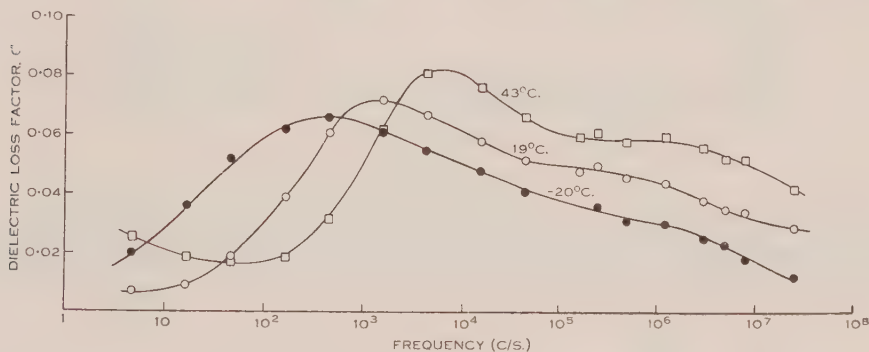


Fig. 1.—Dielectric loss factor of 10-nonadecanol.

a sample of solid 10-nonadecanol, whose thermal history was such that it possessed a relatively low dielectric absorption, was kept at an elevated temperature for several days. It was found that after being in an oven maintained at 62 °C. (4 °C. below the melting point) the dielectric absorption after return to room temperature had increased. The rate of this increase was initially high, for example, in one sample ϵ'' had increased in magnitude by a factor of two at all frequencies after 3 days at 62 °C. but this rate of increase was not maintained after continued storage at 62 °C. Another sample, which was similar in that its thermal history was such that it possessed a relatively low dielectric absorption, was maintained at 55 °C. and the dielectric absorption at room temperature increased by no more than 10 per cent. during 14 days' storage at this temperature.

It would appear then that, provided the sample is not heated to within a few degrees of the melting point, it may be varied up and down in temperature and give, apart from a slow decrease with time, reproducible dielectric properties.

It should be recorded that, although at different times samples of 10-nonadecanol possessing dielectric absorption differing in magnitude by as much as 80 : 1 have been measured in this Laboratory, the absorption curve has always been of approximately the same shape and the absorption has always been recorded in the same frequency region. There is no experimental evidence

to support the suggestion by Meakins and Sack (1951) that as the magnitude of the absorption decreases the absorption moves to higher frequencies. Rather

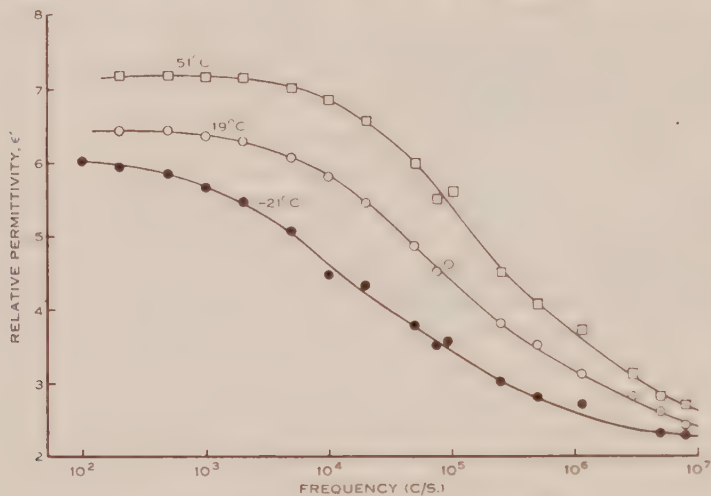


Fig. 2.—Relative permittivity of 14-heptacosanol.

it would appear that for a particular compound the mean relaxation time τ , and the spread of relaxation times has a definite value, or at most the variation from sample to sample is slight. This behaviour would be consistent with the

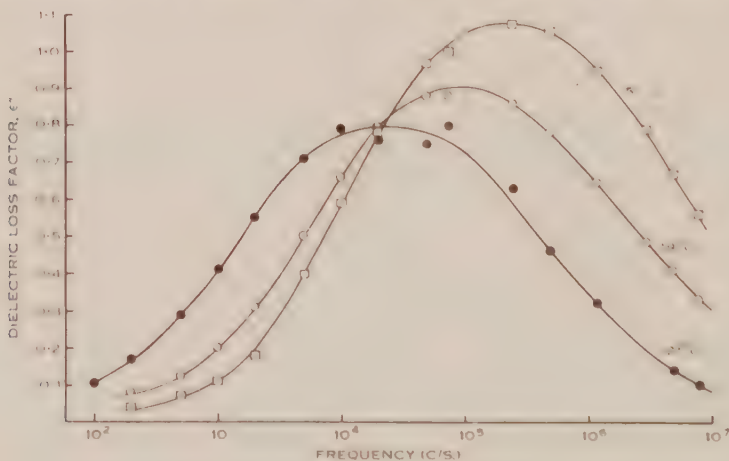


Fig. 3.—Dielectric loss factor of 14-heptacosanol.

dielectric absorption being due to a particular form of the solid which slowly changes with time into another form possessing low dielectric loss.

(b) *Variation of Dielectric Properties of Secondary Alcohols with Temperature*

The loss factor ϵ'' as a function of frequency for three temperatures is shown in Figures 1, 3, and 5 for C_{19} , C_{27} , and C_{35} alcohols respectively. The corresponding variations of ϵ' are shown in Figures 2 and 4 for C_{27} and C_{35} alcohols

respectively. The specimens were heated or cooled to a particular temperature and maintained at that temperature until the dielectric properties had reached a steady value; this always occurred within 1 hour. In all cases the values

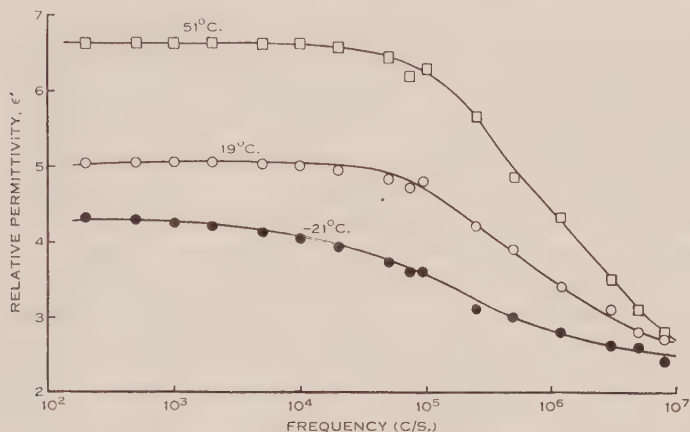


Fig. 4.—Relative permittivity of 18-pentatriacontanol.

of ϵ' and ϵ'' at 19 °C. returned to the same value as before heating or cooling. Values of ϵ' are not shown for 10-nonadecanol since the particular sample of this

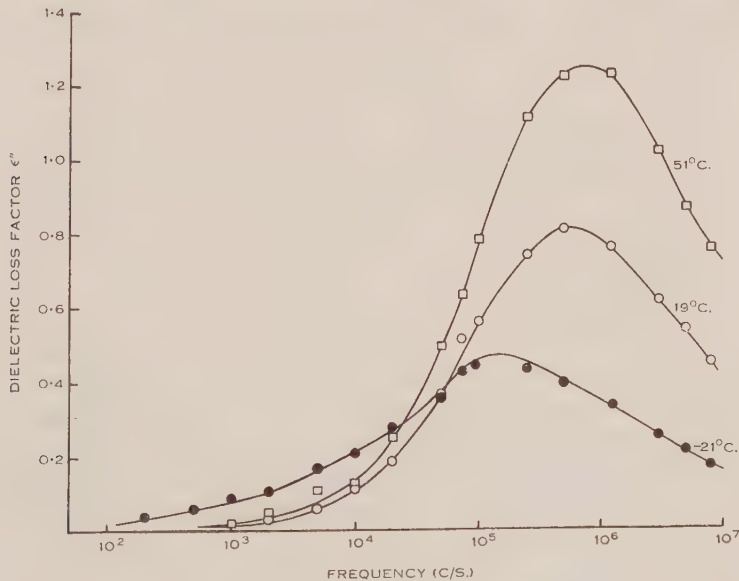


Fig. 5.—Dielectric loss factor of 18-pentatriacontanol.

alcohol used possessed comparatively low absorption and $\epsilon_s - \epsilon_\infty$ was too small to be of assistance in studying variations with temperature.

It can be seen by reference to Figures 1, 3, and 5 that, with increase in temperature, the absorption moves to higher frequencies and increases in

magnitude. Corresponding to this increase in the magnitude of loss factor there is an increase in the quantity $\varepsilon_s - \varepsilon_\infty$. It can be seen that the percentage increase in the magnitude of ε'' is greatest in the compound of highest molecular weight.

Because of the rather flat nature of the curves of ε'' as a function of $\log(\text{frequency})$, there is some uncertainty in assigning a value to the frequency of maximum absorption (f_{max}) at each temperature. However, this is not as great a limitation as might first appear since, in the cases of the C_{27} and C_{35} alcohols, apart from some divergence in one instance on the low frequency side, the same curve fits the measurements at the different temperatures; and consequently the ratios between the different frequencies of maximum absorption can be determined more accurately than individual f_{max} values. An alternative method of arriving at f_{max} is from the frequency at which ε' has the value $\frac{1}{2}(\varepsilon_s + \varepsilon_\infty)$; both methods agree satisfactorily. The values obtained for f_{max} are plotted as a function of $1/T$ in Figure 6. It can be seen that there is an approximately linear relationship between $\log(f_{max})$ and $1/T$ which indicates that the effective mean relaxation time ($\tau = 1/2\pi f_{max}$) can be represented approximately by an expression of the form

$$\tau_n = Ae^{E/kT}, \dots\dots\dots (5)$$

The particular values of A and E are listed in Table 1.

With the C_{19} alcohol, there appears to be, on the basis of this and other experimental observations, two absorption regions which combine to give a

TABLE I

Compound	A (sec.)	E (kcal./mol.)
10-Nonadecanol	10×10^{-10}	6.5
14-Heptacosanol	1.8×10^{-10}	5
18-Pentatriacontanol	1.6×10^{-10}	4.5

very broad absorption. In contrast to the C_{27} and C_{35} alcohols, the shape of the absorption curve alters with temperature. This could arise if there were two absorption regions which had different rates of variation with temperature. The values plotted in Figure 6 are taken from the peak on the low-frequency end and obviously are less certain than those for the other two alcohols.

(c) Primary Alcohols

In Figure 7, ε'' is plotted as a function of frequency for the C_{22} alcohol, with temperature as a parameter. The detailed results are given for one alcohol only since the shape of the absorption curve and the variation in the magnitude of the absorption with temperature are similar in all three alcohols. Like those of the secondary alcohols, the values of ε' and ε'' for each frequency at 19°C. returned to the same values as before heating or cooling.

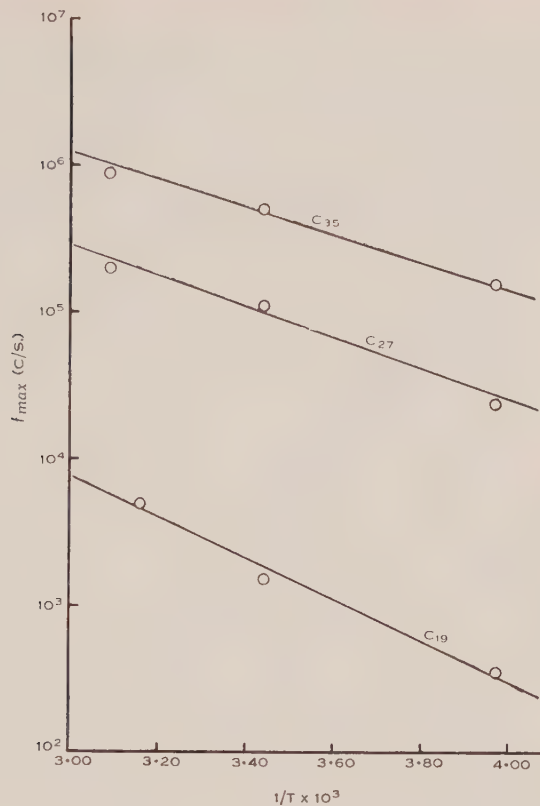


Fig. 6.—Frequency of maximum absorption as a function of the reciprocal of absolute temperature for symmetrical secondary alcohols.

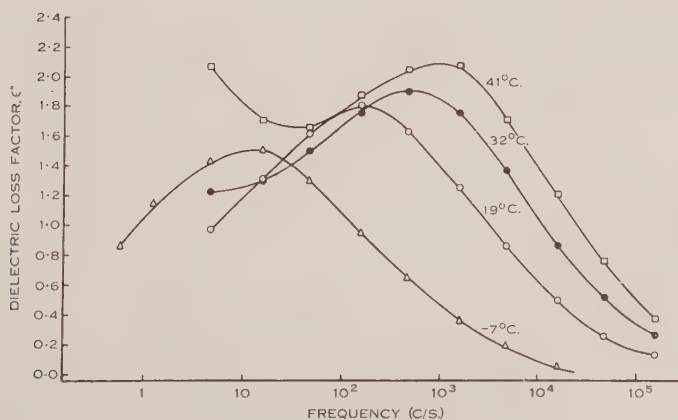


Fig. 7.—Dielectric loss factor of *n*-docosanol,

In the primary alcohols there is at low frequencies a dielectric absorption which is associated with a D.C. conduction and results in ε'' continually increasing with decreasing frequency. This low-frequency effect, unlike the absorption being considered here, is extremely sensitive to the presence of impurities, particularly water (B. V. Hamon and R. J. Meakins, personal communication). As can be seen by reference to Figure 7, the low-frequency effect increases considerably in magnitude with increasing temperature. In the C_{18} sample used this increase was sufficiently great to mask the other absorption at a temperature of 40°C.

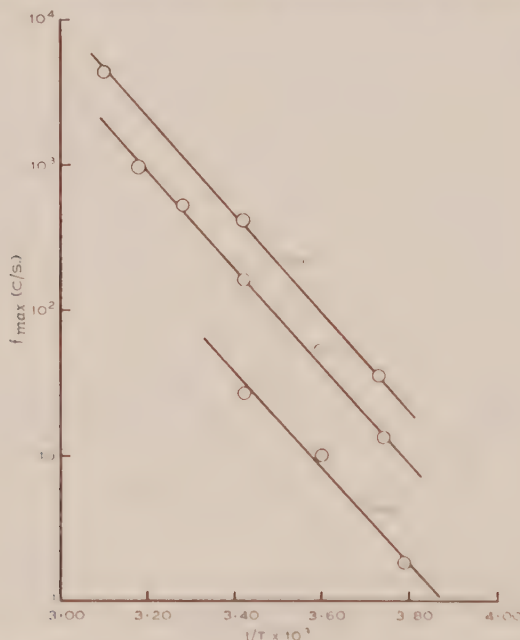


Fig. 8.—Frequency of maximum absorption as a function of the reciprocal of absolute temperature for primary alcohols.

As with the C_{27} and C_{35} secondary alcohols, the shape of the absorption curve does not vary with temperature. In Figure 8 the values of f_{max} are plotted as a function of $1/T$ and the measurements show that an approximate relation exists between the mean relaxation time and temperature of the form of equation

TABLE 2

Compound					A (sec.)
<i>n</i> -Octadecanol	↓	0.8×10^{-15}
<i>n</i> -Docosanol	↓	1.6×10^{-15}
<i>n</i> -Hexacosanol	↓	5.8×10^{-15}

(5). The slopes of the three straight lines in Figure 8 are equal within experimental error, which means that E is the same (15 kcal. mol.) for all three alcohols. Values of A for the different alcohols are set out in Table 2.

V. DISCUSSION

The simplest assumption on which to base the discussion would be that E as determined experimentally is the magnitude of the energy barrier associated with the elementary process, which in this case has been suggested as being either a rotation of a hydroxyl group or a proton transfer. However, if the relaxation time is described by equation (4) or some similar expression, then the quantity E obtained from experimental results will contain a contribution which depends on the energy of interaction between adjacent dipoles. Consequently, E will be spoken of in the discussion as the apparent activation energy.

The fact that the apparent activation energy is independent of molecular weight in the primary alcohols indicates that the absorption in these compounds arises from a mechanism which does not involve the complete molecule. This is consistent with the theory that dielectric absorption in primary alcohols arises from the presence of hydrogen-bonded hydroxyl groups.

The next point to note is that the apparent activation energy differs considerably between the primary and secondary alcohols, being approximately three times as great in the primary alcohols. The non-exponential term A , which includes two factors, the frequency of oscillation of the elementary unit and an entropy term, is also considerably different in the two types of alcohols. These differences could mean that, while the mechanism is the same in the two types of alcohols, there are large differences either in the energy barriers, or in the energy of interaction between neighbouring dipoles, or in both. These differences could arise from differences in the O—O distances between adjacent hydrogen-bonded oxygens or in the environment of the hydroxyl groups. Alternatively, the mechanisms in the primary and secondary alcohols could be different, for example, a proton transfer in one case and a rotation of the hydroxyl group in the other. The apparent activation energy in the secondary alcohols appears to decrease slightly with increasing molecular weight.

It is of interest to compare these apparent activation energies with those associated with dielectric relaxation in ice, which is 13 kcal./mol. (Smyth and Hitchcock 1932), and in water, 5 kcal./mol. (Collic, Hasted, and Ritson 1948). Activation energies, associated with rotation of complete molecules in other aliphatic compounds, can also be compared with those obtained here; the value for methyl stearate ($C_{17}H_{35}COOCH_3$) is 11 kcal./mol. and for methyl behenate ($C_{21}H_{43}COOCH_3$), 17 kcal./mol. (Dryden and Welsh 1951). The similarity in the apparent activation energies in ice and in the primary alcohols may be significant. The primary alcohols crystallize in double molecules, the oxygen atoms from adjacent layers of molecules forming two parallel planes. Consequently, the crystal structure may be such that oxygen atoms form a limited tetrahedral array with three possible positions of the hydrogen atoms along the lines linking adjacent oxygen atoms, each oxygen being covalently bonded to one hydrogen and possessing on the average one hydrogen bond to one of the two possible adjacent hydroxyl groups. Such a structure, which is not inconsistent with what information is available on the crystal structure of primary alcohols (see, for example, Wilson and Ott 1934), could lead to a similar activation energy for dipole orientation as that in ice.

VI. ACKNOWLEDGMENTS

The author wishes to thank Dr. R. J. Meakins for supplying the materials, and colleagues in the Division of Electrotechnology for helpful discussions during the preparation of this paper.

VII. REFERENCES

- COLLIE, C. H., HASTED, J. B., and RITSON, D. M. (1948).—*Proc. Phys. Soc.* **60**: 145.
DRYDEN, J. S., and WELSH, H. K. (1951).—*Aust. J. Sci. Res. A* **4**: 616.
MEAKINS, R. J., and MULLEY, JOAN W. (1951).—*Aust. J. Sci. Res. A* **4**: 365.
MEAKINS, R. J., and SACK, R. A. (1951).—*Aust. J. Sci. Res. A* **4**: 213.
MEAKINS, R. J., and WELSH, H. K. (1951).—*Aust. J. Sci. Res. A* **4**: 359.
SACK, R. A. (1946).—*Trans. Faraday Soc.* **42A**: 61.
SACK, R. A. (1952).—*Aust. J. Sci. Res. A* **5**: 135.
SMYTH, C. P., and HITCHCOCK, C. S. (1932).—*J. Amer. Chem. Soc.* **54**: 4631.
WILSON, D. A., and OTT, E. (1934).—*J. Chem. Phys.* **2**: 231.

DIELECTRIC ABSORPTION AND D.C. CONDUCTIVITY IN *n*-PRIMARY ALCOHOLS

By B. V. HAMON* and R. J. MEAKINS*

[*Manuscript received July 4, 1952*]

Summary

A previous publication described some electrical properties of crystalline *n*-primary alcohols of 16 to 26 carbon atoms. The present paper deals with an extension of the work to the lower homologues down to *n*-octyl alcohol, and includes the results of measurements on both the liquid and crystalline materials. To enable their more extensive investigation, the frequency range of measurement has been extended down to 10^{-4} c/s. by the use of D.C. methods.

The alcohols of the present series resemble the higher homologues in giving two regions of dielectric absorption which overlap to an increasing extent with decreasing molecular weight. The absorption region at lower frequencies merges with the loss due to D.C. conductivity, the latter being much larger than for other long-chain compounds so far investigated.

A feature of the results is the variation in the electrical properties of different samples of the same alcohol. This may be due to variations in the relative proportions of α - and β - crystalline modifications present in different samples.

The results suggest that the absorption and the D.C. conductivity both depend on the presence in the crystals of hydrogen-bond chains. The high frequency absorption appears to be related to that in secondary alcohols and is probably associated with reversal of the hydrogen-bond chains by cooperative rotation of the hydroxyl group about the C-O bonds. The latter mechanism, together with proton transition, could lead to migration of the protons, and this may account for the D.C. conductivity. The absorption at low frequencies is attributed to a Maxwell-Wagner type of mechanism resulting from the situation of the conducting hydrogen-bond chains in a non-conducting matrix.

I. INTRODUCTION

The work described herein forms part of a general programme investigating the effects of hydrogen bonding on the dielectric properties of solids. Some results for *n*-primary alcohols have already been given (Meakins and Mulley 1951), with reference chiefly to the members with 26, 22, and 18 carbon atoms. Preliminary results for the C_{16} compound were also described, but from the shape of the loss-factor curve it was evident that much of the absorption would be found at lower frequencies. This compound and the lower homologues were therefore set aside for separate investigation.

Part of this programme has now been completed, using D.C. techniques (Hamon 1952) to extend the range of measurement down to 10^{-4} c/s. This paper describes the results for five *n*-primary alcohols of chain lengths 8 to 16 carbon atoms.

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

II. MATERIALS

The *n*-primary alcohols were obtained by purification of the commercial products. These were twice vacuum-fractionated, at about 1 mm. pressure, through a 90 cm. column packed with monel metal gauze saddles. The higher members, *n*-hexadecyl and *n*-tetradecyl alcohols, were then recrystallized from methanol to remove oxidative degradation products which may have been formed as a result of their comparatively high distillation temperatures (134 and 120 °C., respectively).

For measurements in the dielectric test cell the final step in the purification involved shaking the molten alcohol with 30 per cent. of its weight of recently activated alumina, and then filtering it through sintered glass directly into the cell. This operation was carried out in a desiccator which was housed in an incubator at a temperature about 10 °C. above the melting point of the alcohol. The inlet tube of the cell was sealed with a rubber stopper to retard the adsorption of atmospheric moisture.

The setting points of the alcohols are given in Table 1.

TABLE 1
SETTING POINTS OF *n*-PRIMARY ALCOHOLS

<i>n</i> -Primary Alcohol	Formula	Setting Point (°C.)
Octyl	$C_7H_{15}.CH_2OH$	-15.7
Decyl	$C_9H_{19}.CH_2OH$	+ 5.5
Dodecyl	$C_{11}H_{23}.CH_2OH$	23.9
Tetradecyl	$C_{13}H_{27}.CH_2OH$	36.6
Hexadecyl	$C_{15}H_{31}.CH_2OH$	48.7

III. APPARATUS AND PROCEDURE

The disk samples for electrical measurement were prepared by melting the alcohol crystals, allowing the melt to solidify, then pulverizing and pressing in a steel mould (Meakins and Sack 1951). With some samples, an additional refinement was the application of a vacuum during melting and solidification of the alcohol. In the subsequent powdering and pressing, exposure to the laboratory atmosphere was reduced to a minimum. For disks of the recrystallized form, referred to in Section IV (a) iv), the crystals were pulverized, then pressed as above.

The disk samples were all measured between mercury electrodes. Measurements over a range of temperature were carried out in the rhodium-plated brass cell shown in Figure 1. The cell requires about 15 cc. of liquid. It was calibrated with dry benzene, and tests showed that surface leakage and dielectric loss in the insulators were negligible.

The temperature of the cell was controlled by suspending it from a cork stopper in a wide-mouthed 2-quart vacuum flask containing sufficient dry light

petroleum or dry kerosene (according to the temperature range) to immerse the cell to about two-thirds of its depth. The contents of the flask were heated by an electrical heater immersed in the liquid, and cooled by solid carbon dioxide placed in a brass tube whose lower end was sealed with a brass foil diaphragm, and whose upper end protruded through the cork. The temperature was measured by thermometers whose bulbs were immersed in the liquid surrounding the cell. The liquid was stirred by bubbles of dry air introduced through a glass tube. It was not difficult to maintain a given temperature to within $\pm 0.5^\circ\text{C}$. at any point in the range from -40 to $+50^\circ\text{C}$. by adjusting the heating rate or the rate at which solid carbon dioxide was added.

Measurements at frequencies from 5 c/s. to 160 kc/s. were made on a permittivity and power-factor bridge designed in this Laboratory (Thompson, unpublished data). Measurements at three lower frequencies (3.9, 1.3, and 0.45 c/s.) were made using a bridge circuit supplied with a sinoidal voltage from a motor-driven potentiometer, and using a Lindemann electrometer as detector. The dielectric loss factors at frequencies lower than 0.1 c/s. were deduced from measurements of the current flowing as a result of applying a sudden D.C. voltage to the sample (Hamon 1952).

Double logarithmic coordinates have been used for the loss factor-frequency curves. With these coordinates a D.C. conductivity corresponds to a straight line with a slope of -45° at the low frequency end of the curve. The conductivity may be calculated from the equation

$$\sigma_{\text{D.C.}} = 0.55 \epsilon'' f \times 10^{-12} \text{ mho cm.}^{-1}, \dots\dots\dots (1)$$

where ϵ'' and f are the loss factor and frequency (c/s.), respectively, at any point on the straight line.

IV. RESULTS AND DISCUSSION

Inconsistencies in the electrical properties of different disks of *n*-hexadecyl alcohol have been noted previously (Meakins and Mulley 1951). The results of a further investigation of these inconsistencies are given in the first part of this Section. The second part (Section IV (b)) gives the dielectric properties of the alcohols measured in the cell (Fig. 1) at a number of temperatures above and below the melting point.

(a) *Investigation of Causes of Variation in Results for Different Specimens of the Same Alcohol*

(i) *Purity*.—It has been shown (Meakins 1949 ; Hamon and Meakins 1950) that low frequency dielectric loss in long-chain compounds can be due to impurities. To see if impurities were responsible for the observed variation in results, a number of different methods of purification and drying were tried with *n*-hexadecyl alcohol. The vacuum-fractionated alcohol was recrystallized up to five times from benzene or methanol. Another method, which had proved very successful in reducing the conductivity of liquid alcohols, was to filter the molten alcohol through recently-activated alumina. For most samples the purified alcohol was dried under a high vacuum over phosphorus pentoxide at

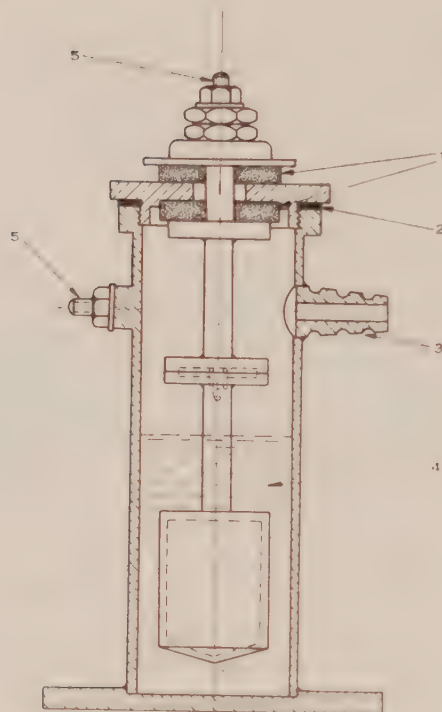
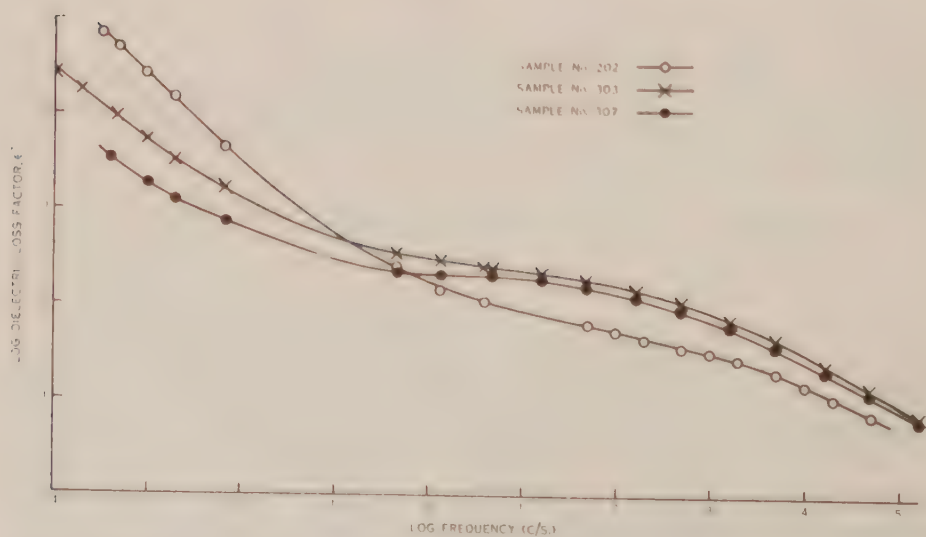


Fig. 1.—Dielectric test cell.

1, polytetrafluorethylene insulators; 2, synthetic rubber gasket; 3, filling tube; 4, sample; 5, terminals.

Fig. 2.—Dielectric absorption of different samples of *n*-hexadecyl alcohol at 20 °C.

temperatures up to 100 °C., although it was found subsequently that heating just above the melting point at 1 mm. pressure gave similar results. Several disks were also pressed under vacuum, using an apparatus with metal bellows designed for this purpose.

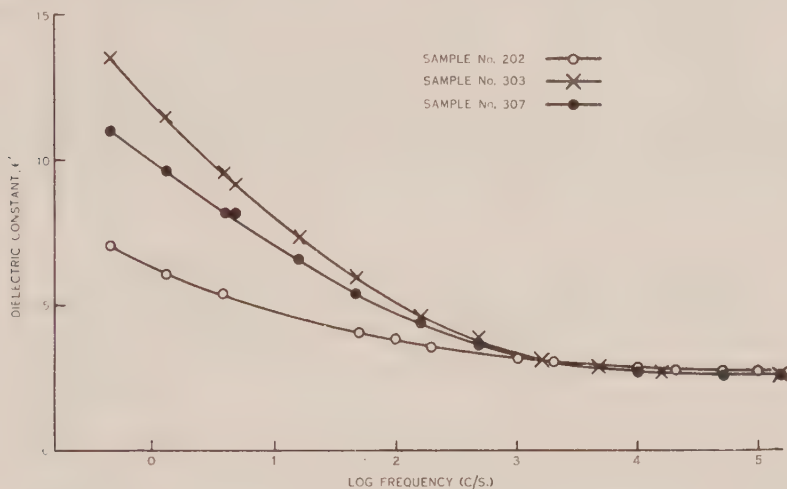


Fig. 3.—Dielectric constant of different samples of *n*-hexadecyl alcohol at 20 °C.

Eleven different samples resulted from this programme. The results for three of these are shown in Figures 2 and 3. The other eight samples gave similar results. If impurities were the main cause of the variation in results, it would

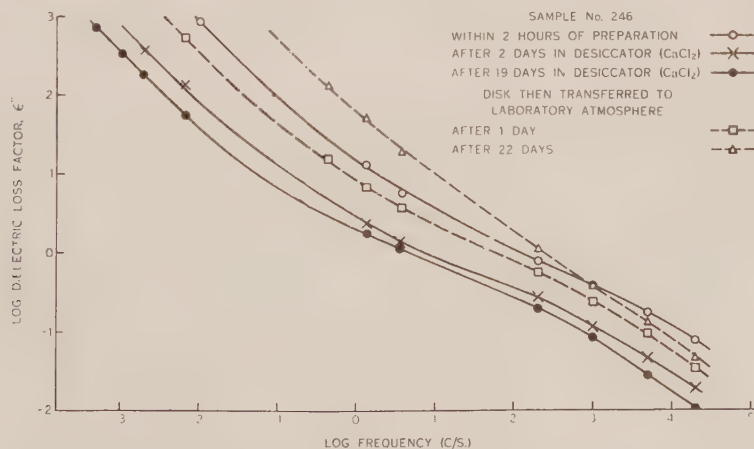


Fig. 4.—Effect of atmospheric moisture on dielectric absorption of *n*-dodecyl alcohol at 20 °C.

be expected that the variation in properties from one sample to another would decrease with continued purification and drying, but this was not observed. A further conclusion from these experiments is that the loss itself cannot be due to impurities,

(ii) *Effect of Adsorbed Moisture.*—In the tests described above, the purification of the hexadecyl alcohol involved careful drying, but some exposure to the atmosphere occurred in preparing the disks. To determine the effect of this exposure, measurements were made on disks that had been stored under controlled humidity conditions.

Figure 4 gives the results of measurements on *n*-dodecyl alcohol. The disk for these experiments was prepared from material that had been dried by filtration through activated alumina. It will be seen that after 2 days in the desiccator, loss factors at frequencies below 1 c/s. are less than one-tenth of their original values, and that a further small decrease occurs after 19 days. At this stage the curve represents, approximately, the dielectric absorption of the pure, dry alcohol.

On exposing the dry disk to the laboratory atmosphere the changes observed above were reversed. The loss factor at frequencies below 1 c/s. increased by a factor of about 10 after 1 day, and by a factor of about 100 after 22 days.

The results given in Figure 4 are complicated by the effects of structural changes (Section IV (b)). Such changes are almost complete after 1 day, however, so that a comparison of the later curves in the diagram should indicate the relative effects of moisture on the high and low frequency absorption. The results show that, as for *n*-dodecyl alcohol, the low frequency absorption is much more strongly affected by moisture.

It is interesting to note that other workers (Meyer and Reid 1933) with *n*-primary alcohols have found that the cooling curves of the compounds with 12 or less carbon atoms are greatly affected by moisture.

Similar tests on a disk of *n*-hexadecyl alcohol showed very little change due to storage in a desiccator. To confirm this, further experiments were carried out on disks made from crystals that had been exposed to 90 per cent. relative humidity before pressing.

Figure 5 shows the results of these tests. Three disks were made from the one batch of material which had been recrystallized from methanol. The first disk was made after drying the crystals over P_2O_5 , and the second and third after exposure for 2 and 22 days, respectively, to 90 per cent. relative humidity. The results show that the ingress of moisture into crystalline *n*-hexadecyl alcohol is slow, and that consequently the short exposure that occurs in preparing a disk should have a negligible effect on its properties. As is usual, the effect of the adsorbed moisture is greater at the lower frequencies. The effect of atmospheric moisture on *n*-tetradecyl alcohol is similar to that shown in Figure 5 for *n*-hexadecyl alcohol.

(iii) *Varying Particle Size.*—Measurements were made with disks prepared from *n*-hexadecyl alcohol of widely varying particle size. The results were within the range of those given in Figures 2 and 3, indicating that particle size has little effect on the dielectric properties.

(iv) *Rate of Crystallization from the Melt.*—A disk prepared from *n*-hexadecyl alcohol which had been melted and then rapidly chilled gave results similar to those in Figures 2 and 3.

(v) *Presence of Different Crystalline Modifications.*—From the results of these tests, it appears unlikely that variations in the dielectric properties of different specimens of *n*-hexadecyl alcohol can be accounted for by the presence of impurities or by slight differences in procedure. It is more likely that such variations are due to the presence of varying amounts of two different crystalline phases, each phase having its own characteristic dielectric properties.

This explanation is supported by X-ray data obtained by Wilson and Ott (1934) who found that *n*-hexadecyl alcohol crystallizes from the melt as a mixture of α - and β -modifications which differ in the angle of inclination of the molecules to the base of the unit cell. The relative proportions of the two modifications

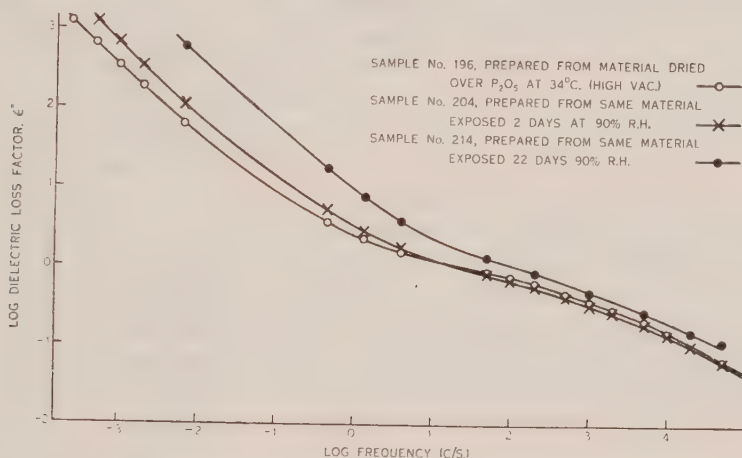


Fig. 5.—Effect of atmospheric moisture on dielectric absorption of *n*-hexadecyl alcohol at $20^\circ C$.

showed a wide variation from one specimen to another. Other evidence of the existence of two crystalline modifications of *n*-primary alcohols has been obtained from cooling-curve studies (Meyer and Reid 1933).

The position is not simple, however, since disks made directly from recrystallized alcohol gave results similar to disks made from alcohol that had first been melted, except that the absorption above 1 c/s. was somewhat less. From the work of Kolp and Lutton (1951) it was expected that the recrystallized material would be composed almost wholly of the β -modification.

(b) Variation of Dielectric Properties with Temperature

(i) C_{16} , C_{14} , and C_{12} Alcohols.—Some loss-factor and dielectric constant curves for *n*-hexadecyl alcohol, measured in the dielectric test cell, are given in Figures 6 and 7. The curves for the sample at room temperature are similar to previous results for disk samples (see Fig. 2), except that the former gives somewhat larger absorption at frequencies above 1 c/s. Similar differences are obtained with cell and disk samples of the C_{14} and C_{12} alcohols. They have been found to be due to crystal orientation, which will be discussed in more detail in a later paper. Presumably, the low frequency absorption would also

be affected, but the differences would be masked by the relatively large variations for different samples at these frequencies.

It is important to note that the dielectric absorption of the cell samples is not less than that of the disks. Since the former were prepared and measured

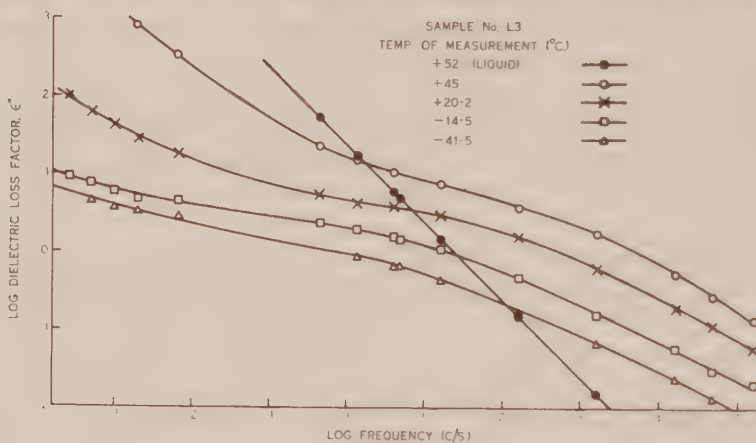


Fig. 6.—Dielectric absorption of *n*-hexadecyl alcohol measured in closed cell.

under very dry conditions, this result provides further evidence that the observed dielectric absorption is a property of the alcohol itself and is not due to the presence of traces of water.

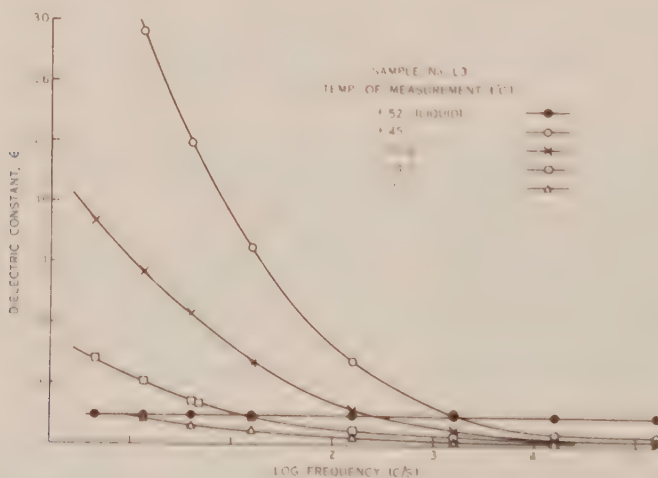


Fig. 7.—Dielectric constants of *n*-hexadecyl alcohol measured in closed cell.

The four loss factor curves for the crystalline alcohol cover a temperature range of 86.5° C. There is very little change of shape at the higher frequencies and in this region the loss factor increases by about 10 per cent. for each degree rise in temperature.

At frequencies below 1 c/s., however, the variation with temperature is more complex and appears to depend considerably on the proximity to the melting point. There is comparatively little difference between the results at -41.5 and -14.5 °C., and the shape of the curves at the lowest frequencies suggests that the D.C. conductivity at these temperatures is very small ($\ll 1 \times 10^{-15}$ mho cm.⁻¹). A comparison of these results with those at 20.2 °C. ($\sigma_{D.C.} < 1 \times 10^{-14}$ mho cm.⁻¹) and 45 °C. ($\sigma_{D.C.} = 1 \times 10^{-12}$ mho cm.⁻¹) indicates that the low frequency absorption and D.C. conductivity increase more and more rapidly as the melting point is approached.

Samples of *n*-hexadecyl and *n*-tetradecyl alcohols stored in the test cell at room temperature showed slight decreases in dielectric loss, the amount of the decrease being about 10 per cent. in 24 hours. With *n*-dodecyl alcohol the effect was larger, the absorption decreasing within 24 hours to about half its original value at frequencies below about 1 c/s. and to about one-third at higher frequencies. Similar changes have been observed in the secondary alcohols (Meakins and Sack 1951) and are considered to be due to changes in crystal structure.

The loss-factor curve for liquid *n*-hexadecyl alcohol, shown in Figure 6, approximates to a straight line, making an angle of 45° with the abscissa. This indicates that the dielectric loss in the liquid alcohol is due almost entirely to D.C. conduction and, in agreement with this, the dielectric constant shows no increase with decreasing frequency. The dielectric constant value of 3.9 is appreciably larger than for the solids at high frequencies (about 2.5), indicating the presence of considerable absorption at frequencies above the present range of measurement.

Crystalline *n*-tetradecyl alcohol, measured in the dielectric cell, gave results similar to *n*-hexadecyl alcohol. *n*-Dodecyl alcohol is similar at low frequencies, but the loss-factor curve falls off much more steeply at frequencies above about 1 c/s.

(ii) *C*₁₀ and *C*₈ Alcohols.—The *C*₁₀ and *C*₈ compounds are liquids at room temperature and can therefore be compared in the solid state only at lower temperatures. Some results for *n*-octyl alcohol are given in Figure 8. At -42.6 °C., its dielectric absorption above 1 c/s. is similar to that of *n*-hexadecyl alcohol, but with decreasing frequency it increases much more rapidly. At the lowest frequency of measurement the loss factor is about 30 times larger than for *n*-hexadecyl alcohol at the same frequency and temperature, and the shape of the curves suggests that the difference in the D.C. conductivities of the two compounds is even larger.

It is interesting to note that the loss-factor curves given in Figure 8 for crystalline *n*-octyl alcohol show no evidence of the maximum at 305 c/s. (-23.6 °C.) reported for this compound by Dalbert, Magat, and Surdut (1949).

The results for the liquid are similar to those described above for the *C*₁₆ compound and are typical of all five alcohols of the present series. The observed dielectric loss is due almost entirely to D.C. conductivity, which increases considerably with decreasing molecular weight. Thus, the value of the D.C.

conductivity determined (see equation (1)) from the curve for *n*-octyl alcohol at 20 °C. ($3 \cdot 10^{-9}$ mho cm.⁻¹) is 300 times larger than that for *n*-hexadecyl alcohol at 52 °C. (1×10^{-11} mho cm.⁻¹).

The loss-factor curve for liquid *n*-octyl alcohol at -6.5 °C. (see Fig. 8) shows a minimum at about 50 kc/s. Further measurements were made with a *Q*-meter up to 10 Mc/s. and it was found that the loss factor increases with frequency over the whole of this range. This is evidently the beginning of the absorption region reported by other authors (Dalbert, Magat, and Surdut 1949) to give a maximum at about 30 Mc/s. at 1.5 °C. Similar absorption was observed in all

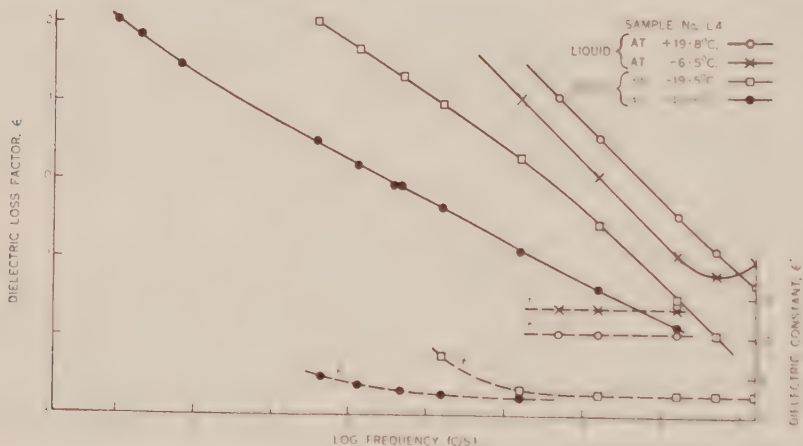


Fig. 8.—Dielectric properties of *n*-octyl alcohol.

the liquid *n*-primary alcohols of the present series and was accompanied in each case by an enhancement of the dielectric constant towards lower frequencies. The values of the dielectric constant obtained for liquid *n*-octyl alcohol are in good agreement with those given by Dalbert, Magat, and Surdut (1949).

V. GENERAL DISCUSSION OF RESULTS

(a) Crystalline *n*-Primary Alcohols

The five alcohols from C_8 to C_{16} discussed in the present paper are similar to the higher homologues (Meakins and Mulley 1951) in giving very large dielectric absorption over a wide range of frequencies. They also give evidence of separate absorption regions, but with more overlapping than in the higher homologues.

In the following discussion, the term "low frequency absorption" refers to the region of dielectric loss below about 1 c/s. but excluding the loss due to D.C. conductivity. The loss above this frequency is referred to as the "high frequency absorption". With the C_{16} and C_{14} compounds the two regions are separated by a marked change of slope of the loss-factor curves but this is less evident in the lower homologues. There are, however, differences in behaviour of the absorption at high and low frequencies, as, for example, towards temperature variation, and the ingress of moisture. This suggests that two different mechanisms are involved. In previous work (Meakins and Mulley

1951), the predominant factor affecting the dielectric properties has been assumed to be the presence in the crystals of hydrogen-bonded chains of hydroxyl groups. It is necessary to show how these chains can function in different mechanisms to account for the observed electrical properties.

(i) *The High Frequency Absorption Region.*—A characteristic feature observed for this absorption is its displacement to lower frequencies with decreasing molecular weight. In this it resembles the absorption in secondary alcohols (Meakins and Sack 1951) and can be explained by a similar mechanism involving hydrogen-bonded chains of hydroxyl groups which form giant dipoles capable of reversal by cooperative rotation of the individual hydroxyl groups about the C—O bonds.

The evidence for this mechanism has been discussed in the previous publications on secondary and primary alcohols. The main points are :

- (1) The absorption is not displaced to lower frequencies with increasing molecular weight and therefore does not depend on the rotation of complete molecules.
- (2) The absorption is much larger than in any other long-chain compounds so far investigated, although the dipole moment of individual hydroxyl groups is not large (μ = about 1.6). This suggests that the dipoles are grouped together to form units of large electric moment.
- (3) The dielectric absorption in alcohols occurs at much lower frequencies than would be expected to result from the orientation of individual hydroxyl groups. This would be expected from a cooperative effect.
- (4) Mixing the alcohols with hydrocarbons to form dilute solid solutions eliminates the dielectric absorption, indicating that the latter is due to groups, and not to individual dipoles.
- (5) A consideration of the probable molecular arrangement in crystalline long-chain alcohols suggests that the most likely grouping of the dipoles is a hydrogen bonding of the hydroxyl groups to form chains parallel to one of the axes of the unit cell.

(ii) *D.C. Conductivity in n-Primary Alcohols.*—In addition to the mechanism of rotation mentioned above, it may also be possible for protons to transfer from one oxygen atom to another. A combined mechanism involving successive rotation and transfer should then result in a migration of protons through the material (Stearn and Eyring 1937) and can account for the D.C. conductivity of the *n*-primary alcohols. A similar combined mechanism has already been suggested to occur in ice (Pauling 1945) in which the hydrogen-bond system should be somewhat similar to that in the double layer of hydroxyl groups of the *n*-primary alcohols (Wilson and Ott 1934).

(iii) *The Low Frequency Absorption in n-Primary Alcohols.*—An important feature of the low frequency absorption region is that it always varies in a similar manner to the D.C. conductivity. This suggests that the mechanisms in the two cases may be somewhat similar. A possible explanation is that the conducting hydrogen-bond chains, such as those proposed above to account for the D.C. conductivity, may cause low frequency absorption when they do not

pass completely through the material. In this case, they would appear in the system as conductors surrounded by a non-conducting matrix and would thus form a Maxwell-Wagner system, which would be expected to give rise to dielectric polarization. In a simple Maxwell-Wagner system with spherical conducting particles dispersed in a non-conducting matrix, the absorption appears as a Debye-type curve with a single relaxation time. With conductors of variable length, such as the hydrogen-bond chains in crystalline primary alcohols, a range of relaxation times would be expected.

It should be mentioned that recent measurements in these Laboratories have shown the presence of similar low frequency absorption in a variety of slightly conducting solids which, however, do not contain hydroxyl groups. It is probable that a Maxwell-Wagner system resulting from macroscopic inhomogeneities is responsible for the absorption in all these materials, although the mechanism of conduction within each conducting region is certainly not the same as that proposed above.

(b) Liquid *n*-Primary Alcohols

Dielectric absorption in the liquid alcohols occurs at frequencies above 200 kc/s., and has been discussed for *n*-octyl alcohol by Dalbert, Magat, and Surdut (1949). These authors consider the absorption to be related to the duration of life of molecular aggregates, but do not specify the nature of the aggregates, or discuss the way in which their duration of life could produce dielectric loss. An alternative explanation is that the absorption is due to a hydrogen-bonded chain mechanism such as that discussed above for crystalline alcohols. Hydrogen-bonded chains in a liquid alcohol would be much shorter than in the crystal, due to the smaller extent of the molecular order, and this would account for the relatively high frequency of the absorption.

It is interesting to note that the presence of hydrogen-bonded chains in liquid alcohols has also been inferred from infra red data (Mecke 1950).

VI. ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance of Miss M. C. Clark, Division of Electrotechnology, C.S.I.R.O., in making many of the electrical measurements in this work.

VII. REFERENCES

- DALBERT, M., MAGAT, M., and SURDUT, A. (1949). "Polarisation de la Matière" p. 14. (Colloques Internationaux: Paris.)
HAMON, B. V. (1952).—*Instn. Elect. Engrs. Monogr.* No. 27.
HAMON, B. V., and MEAKINS, R. J. (1950).—*Nature* **166**: 29.
KOLP, D. G., and LUTTON, E. S. (1951).—*J. Amer. Chem. Soc.* **73**: 5593.
MEAKINS, R. J. (1949).—*Aust. J. Sci. Res. A* **2**: 405.
MEAKINS, R. J., and MULLEY, JOAN W. (1951).—*Aust. J. Sci. Res. A* **4**: 365.
MEAKINS, R. J., and SACK, R. A. (1951).—*Aust. J. Sci. Res. A* **4**: 213.
MECKE, R. (1950).—*Discuss. Faraday Soc.* No. 9: 161.
MEYER, J. D., and REID, E. E. (1933).—*J. Amer. Chem. Soc.* **55**: 1574.
PAULING, L. (1945). "The Nature of the Chemical Bond." p. 302. (Cornell Univ. Press.)
STEARNS, A. E., and EYRING, H. (1937).—*J. Chem. Phys.* **5**: 113.
WILSON, D. A., and OTT, E. (1934).—*J. Chem. Phys.* **2**: 231.

THE KINETICS OF ADSORPTION AT LIQUID SURFACES

By K. L. SUTHERLAND*

[Manuscript received April 30, 1952]

Summary

It is shown that convection currents govern the time rate of fall of surface tension of aqueous solutions when measurements are made in the usual way. The complete interpretation of such data is not possible at present so that an experimental method is suggested whereby data may be obtained which could be interpreted. The mathematical equations are given for this method.

I. INTRODUCTION

The composition of the surface of a solution usually differs from that of the bulk, and may be calculated from the surface tension using the Gibbs adsorption isotherm. A new surface formed on a solution will initially have a composition the same as that of the bulk. To reach equilibrium solute will move from the body of the solution to this new surface. During this period the surface tension will fall to the equilibrium value. Even for aqueous solutions of simple organic compounds, like the alcohols, the factors governing the rate at which equilibrium is reached are not yet known.

Isaacs and Sutherland (1952) used oscillating jets to determine the time rates of fall in surface tension for solutions of *n*-heptanol and 3-methyl-1-butanol. They were able to show that the rates were dependent upon the orifice used to produce the jet. As a consequence, they deduced that transport of solute to a surface was governed, *inter alia*, by liquid movements within the jet and hence the rate of change in surface tension was determined by the hydrodynamics of the measuring system. A similar behaviour must arise in all those experimental methods which produce a flow of liquid in the system. On the other hand methods such as the Wilhelmy plate method are free from this objection. Data obtained by this method (Addison and Hutchinson 1949) are therefore examined. Such data may enable us to determine if there is an energy barrier to the movement of solute to and from a surface or whether diffusion is the sole factor determining the fall in surface tension.

Guaralla and Mibishan-Saraga (1949) and Saraga (1950) have examined the desorption of a solute from a surface under conditions which should give clear evidence as to mechanism. They found that their solutions were stirred by convection but claimed that their results were consistent with a process in which only diffusion was important.

The effect of stirring by convection will be considered in the present paper and it will be shown that the rate of fall of surface tension depends upon the

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

dimensions of certain parts of the equipment which have not been considered relevant by those making investigations on this problem. From these considerations, criteria for future experimental measurements will be presented together with the necessary mathematical equations to determine the influence of diffusion, and of any energy barrier on the rate of fall of surface tension.

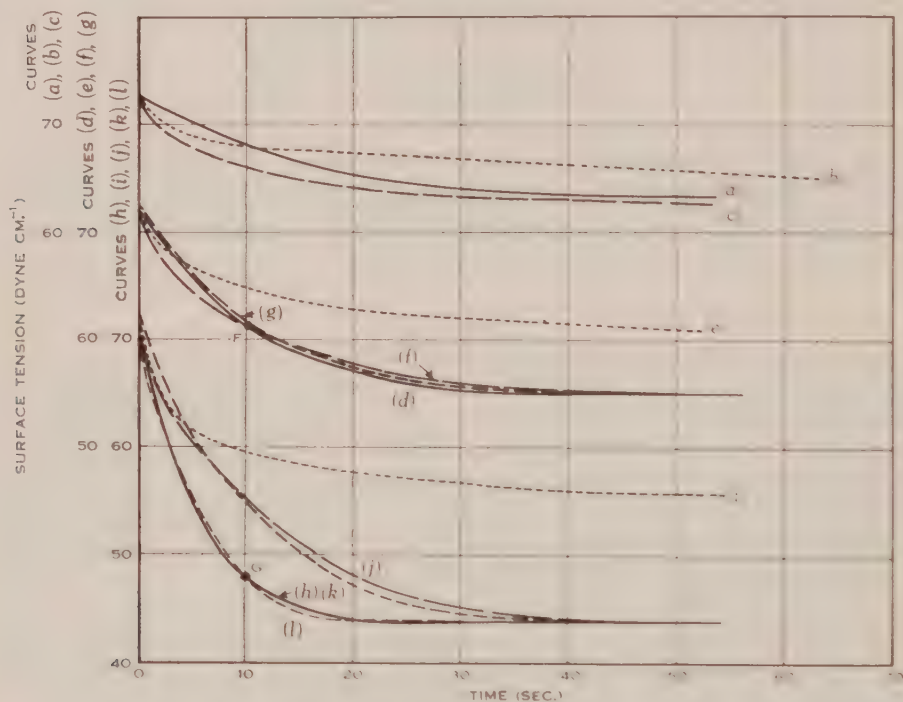


Fig. 1.—The time rate of fall of surface tension of solution of *n*-decanol. Temperature 20 °C. Experimental curves by Addison and Hutchinson (1949).

Curve (a) 0.0005 per cent., experimental; curve (b) 0.0005 per cent., theoretical, unstirred, (eqn. (5)); curve (c) 0.0005 per cent., theoretical, stirred, $2l = 5.96 \times 10^{-3}$ cm. (eqn. (23)); curve (d) 0.0010 per cent., experimental; curve (e) 0.0010 per cent., theoretical, unstirred, (eqn. (5)); curve (f) 0.0010 per cent., theoretical, stirred, $2l = 5.96 \times 10^{-3}$ cm. (eqn. (23)); curve (g) 0.0010 per cent., theoretical, stirred, $E = 15.1RT$ (eqn. (35)); curve (h) 0.0020 per cent., experimental; curve (i) 0.0020 per cent., theoretical, unstirred, (eqn. (5)); curve (j) 0.0020 per cent., theoretical, stirred, $2l = 5.96 \times 10^{-3}$ cm. (eqn. (23)); curve (k) 0.0020 per cent., theoretical, stirred, $E = 15.1RT$ (eqn. (35)); curve (l) 0.0020 per cent., theoretical, stirred, $2l = 2.04 \times 10^{-3}$ (eqn. (23)).

Point F was used to obtain a value of $2l$ or E and thus produce curves (c), (f), (g), (j), (k). Point G was used to obtain $2l$ to produce curve (l).

II. DATA OF ADDISON AND HUTCHINSON

The solution (*n*-decanol in water) used by these authors was contained in a rectangular trough of depth (probably) greater than 1 cm. The surface was "swept" clean with a barrier and the fall of surface tension with time recorded using the Wilhelmy plate method. This method involves but little disturbance of the surface of the solution. Their results are presented in Figure 1.

The observed rate of fall of surface tension may arise in the following five ways:

- (a) The solute diffuses from unstirred solution to the surface.
- (b) The solute diffuses from unstirred solution to the surface. Before entering or leaving the surface solute must surmount an energy barrier.
- (c) The liquid deep in the trough is thoroughly stirred but near the surface it is quiescent. Solute diffuses through the unstirred layer to the surface.
- (d) The transfer is similar to that in (c) but there is an energy barrier at the surface hindering the movement of solute molecules to and from the surface.
- (e) The liquid is everywhere well stirred and the sole controlling factor is an energy barrier to movement to and from that surface.

Stirring in systems (c), (d), and (e) may arise through convection currents and the significant features of these currents will be discussed later.

The diffusion of solute to and from the surface in each of the above cases may be described mathematically. The appropriate differential equations will be given followed by their solutions. In all systems an infinite body of liquid is bounded by a plane, the surface of the solution. The following symbols are used:

c = concentration,

c_0 = initial concentration,

c_e = concentration in equilibrium with a surface of given surface tension,

c_s = surface concentration (see text),

G = free energy of transfer of solute from solution to surface,

E = free energy barrier to transfer of solute to and from the surface,

x = depth of solution from an appropriate plane,

\bar{M} = average value of the ratio Γ/c_s and is defined by (Rideal and Sutherland 1952)

$$\frac{1}{c_e} \int_0^{c_e} \frac{\Gamma}{c} dc,$$

Γ = surface excess of solute at any time, t ,

Γ_s = surface excess at equilibrium,

D = diffusion constant,

R = gas constant,

T = absolute temperature,

$2l$ = thickness of unstirred solution,

t = time,

Z_s = frequency factor for collision at the surface (taken as $\sqrt{RT/2\pi m}$, where m = molecular weight of the solute).

(i) *System (a): Unstirred Solution, Surface at $x=0$*

It will be assumed throughout that Fick's law is obeyed. Consequently diffusion within the solution is given by

$$D \frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t} \quad x > 0, \quad t > 0. \quad \dots\dots\dots (1)$$

At the surface the flux of solute from solution must be equal to the gain in surface excess, that is,

$$D \frac{\partial c}{\partial x} - \frac{d\Gamma}{dt} = 0, \quad x=0, \quad t>0.$$

At present it is only possible to obtain solutions to these differential equations if Γ is a linear function of the concentration. This is not true but over a sufficiently narrow range of concentration, we may write

$$\Gamma = \bar{M}c,$$

where \bar{M} is an average defined above and is to be considered as independent of concentration. The boundary condition is then

$$D \frac{\partial c}{\partial x} - \bar{M} \frac{dc}{dt} = 0, \quad x=0, \quad t>0. \quad \dots\dots\dots (2)$$

In addition there are the boundary conditions

$$c=0, \quad x=0, \quad t=0, \quad \dots\dots\dots (3)$$

$$c=c_0, \quad x>0, \quad t=0. \quad \dots\dots\dots (4)$$

The solution is

$$\left(\frac{c}{c_0}\right)_{x=0} = \frac{\Gamma}{\Gamma_f} = 1 - e^{Dt/\bar{M}^2} \operatorname{erfc} \frac{\sqrt{Dt}}{\bar{M}}, \quad \dots\dots\dots (5)$$

and

$$\operatorname{erfc} \frac{\sqrt{Dt}}{\bar{M}} = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{e^{-\zeta^2} d\zeta}{\sqrt{Dt/\bar{M}}}, \quad \dots\dots\dots (6)$$

A more elegant method of dealing with this case is presented by Ward and Tordai (1946). They avoid the use of the mean, \bar{M} , but the solution is cumbersome in application.

(ii) System (b): *Unstirred Solution, Energy Barrier*

As boundary conditions we have that the flux of solute to the surface must be equal to the increase in amount in the surface (equation (8)) and must also equal the rate of passage across the potential barrier. The net rate is given by collision frequency, concentrations of molecules in the surface and solution just underlying the surface, and by the magnitude of the free energy barrier. This leads to equation (9).

$$D \frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t}, \quad x>0, \quad t>0, \quad \dots\dots\dots (7)$$

$$D \frac{\partial c}{\partial x} - \bar{M} \frac{dc_s}{dt} = 0, \quad x=0, \quad t>0, \quad \dots\dots\dots (8)$$

$$D \frac{\partial c}{\partial x} + Z_f e^{-E/RT} (c_s - c) = 0, \quad x=0, \quad t>0, \quad \dots\dots\dots (9)$$

$$c=c_0, \quad x>0, \quad t=0, \quad \dots\dots\dots (10)$$

$$c_s=0, \quad x=0, \quad t=0, \quad \dots\dots\dots (11)$$

where c_s is defined by

$$c_s = \frac{\Gamma}{\delta} e^{-G/RT}, \quad \dots \quad (12)$$

where δ is the "thickness" of the adsorbed layer. The quantity $(1/\delta) \exp(-G/RT)$ may be obtained directly from the equilibrium curve of surface tension against concentration.

The solution is

$$\frac{c}{c_0} = 1 - \frac{h}{\beta - \alpha} \left[e^{\alpha x + Dt\alpha^2} \operatorname{erfc} \left\{ \frac{x}{2\sqrt{Dt}} + \alpha\sqrt{Dt} \right\} - e^{\beta x + Dt\beta^2} \operatorname{erfc} \left\{ \frac{x}{2\sqrt{Dt}} + \beta\sqrt{Dt} \right\} \right], \quad \dots \quad (13)$$

where

$$\left. \begin{aligned} \alpha &= \frac{h}{2} - \sqrt{\frac{h^2}{4} - \frac{h}{MD}}, \\ \beta &= \frac{h}{2} + \sqrt{\frac{h^2}{4} - \frac{h}{MD}}, \\ h &= Z_f e^{-E/RT} / D. \end{aligned} \right\} \quad \dots \quad (14)$$

At the surface ($x=0$)

$$\frac{\Gamma}{\Gamma_f} = 1 - \frac{\beta}{\beta - \alpha} e^{Dt\alpha^2} \operatorname{erfc}(\alpha\sqrt{Dt}) + \frac{\alpha}{\beta - \alpha} e^{Dt\beta^2} \operatorname{erfc}(\beta\sqrt{Dt}). \quad \dots \quad (15)$$

(iii) *System (c): Stirred Solution, Unstirred Layer 21, $x=0$ at Centre of Unstirred Layer*

$$D \frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t}, \quad -l < x < l, \quad t > 0, \quad \dots \quad (16)$$

$$c = c_0, \quad -l < x < l, \quad t = 0, \quad \dots \quad (17)$$

$$c_s = 0, \quad x = l, \quad t = 0, \quad \dots \quad (18)$$

$$c = c_0, \quad x = -l, \quad t \geq 0, \quad \dots \quad (19)$$

$$D \frac{\partial c}{\partial x} + \bar{M} \frac{dc_s}{dt} = 0, \quad x = l, \quad t > 0. \quad \dots \quad (20)$$

The solution is

$$\frac{c}{c_0} = 1 - \sum_{n=1}^{\infty} \frac{2\alpha_n m e^{-\kappa\alpha_n^2 t} \sin \alpha_n(y+1)}{4\alpha_n m \sin 2\alpha_n + (2\alpha_n^2 m - 1) \cos 2\alpha_n}, \quad \dots \quad (21)$$

where $y = x/l$, $\kappa = D/l^2$, $m = \bar{M}/l$, and α_n are the roots of

$$\alpha \tan 2\alpha = \frac{1}{m}. \quad \dots \quad (22)$$

At the surface ($y=1$, $\frac{c}{c_0} = \frac{\Gamma}{\Gamma_f}$)

$$\frac{\Gamma}{\Gamma_f} = 1 - \sum_{n=1}^{\infty} \frac{e^{-\kappa\alpha_n^2 t}}{m\alpha_n^2 + 0.5 + \frac{1}{m}}. \quad \dots \quad (23)$$

(iv) *System (d): Stirred Solution, Unstirred Layer 2l, Energy Barrier to Transfer at the Surface $x=1$*

$$D \frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t}, \quad -l < x < l, \quad t > 0, \quad \dots \quad (24)$$

$$c = c_0, \quad -l < x < l, \quad t = 0, \quad \dots \quad (25)$$

$$c_s = 0, \quad x = l, \quad t = 0, \quad \dots \quad (26)$$

$$c = c_0, \quad x = -l, \quad t \geq 0, \quad \dots \quad (27)$$

$$D \frac{\partial c}{\partial x} + \bar{M} \frac{dc_s}{dt} = 0, \quad x = l, \quad t > 0, \quad \dots \quad (28)$$

$$D \frac{\partial c}{\partial x} + H(c - c_s) = 0, \quad x = l, \quad t > 0, \quad \dots \quad (29)$$

where

$$H = Z_f e^{-E/RT}. \quad \dots \quad (30)$$

The surface excess is given by

$$\frac{\Gamma}{\Gamma_f} = 1 - \sum_n \frac{e^{-zx_n^2 t}}{m x_n [x_n \tan^2 2x_n + \frac{1}{2} \tan 2x_n + x_n (h - 1)]}. \quad \dots \quad (31)$$

where z , m are as previously defined, $h = zH$, and x_n are the roots of

$$\alpha^2 m h + \alpha m \tan 2\alpha - 1 = 0. \quad \dots \quad (32)$$

(v) *System (e): Complete Stirring, Energy Barrier at Surface*

$$\frac{d\Gamma}{dt} = Z_f e^{-E/RT} (c_0 - c_s), \quad t > 0, \quad \dots \quad (33)$$

$$c_s = 0 \quad t = 0. \quad \dots \quad (34)$$

The solution is

$$\frac{\Gamma}{\Gamma_f} = 1 - \exp \left(- \frac{Z_f e^{-E/RT} t}{\bar{M}} \right), \quad \dots \quad (35)$$

where \bar{M} is as previously defined. This may also be solved explicitly with \bar{M} as a function of concentration.

(vi) *Numerical Calculations*

The diffusion constant, D , was obtained from the Einstein-Smoluchowski formula ($5 \cdot 1 \cdot 10^{-6}$ cm.² sec.⁻¹). The value of \bar{M} is shown in Figure 2. It is not constant particularly for solutions near saturation and these solutions were not examined. The fall in surface tension predicted by equation (5) is slower than the experimental rates (Fig. 1). Equation (15) is not plotted for it predicts even slower changes of surface tension than equation (5) because there is an energy barrier hindering the entry of molecules into the surface. All the remaining processes (equations (23), (31), and (35)) give predicted rates in reasonable agreement with the experiments. Equation (23) has one adjustable constant (the thickness of the unstirred layer $2l$), equation (35) one adjustable constant (the free energy of activation E), and equation (31) has two adjustable

constants ($2l$, E). Only one experimental point (see Fig. 1, point F) was used to determine a constant (equations (23) or (35)) and thus to determine the three curves. For the highest concentration a much better fit may be achieved by choosing one point (e.g. point G) on the experimental curve and using this to find $2l$ (curve I) or E for this curve. Using two points on the curve to give two constants (equation (31)) a still better fit is achieved but it does not seem necessary at this stage to use a relation involving two constants.

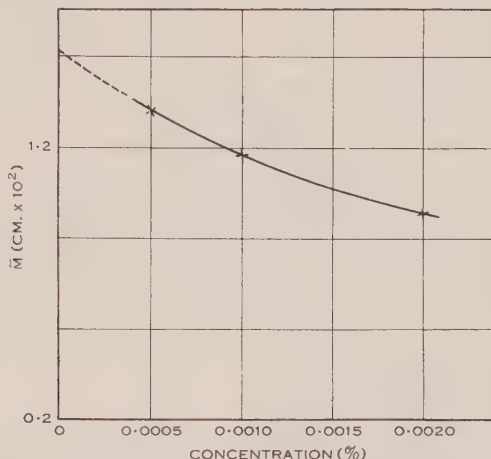


Fig. 2.—The variation of the coefficient \bar{M} with concentration of decanol.

III. DATA OF GUASTALLA AND SARAGA

These data describe the desorption of lauric acid from a monolayer kept at constant pressure and are presented in Figure 3. Guastalla and Mibashan-Saraga (1949) deduced that there must be convection in the liquid because the rate of desorption became constant after some time. They determined the thickness of an unstirred layer from the steady state. Saraga (1950) has attempted to analyse the non-steady state but her analysis is incomplete.

The steady state indicates that systems (a) and (b) described in Section II cannot apply. The presence of a non-steady state excludes system (e) since the loss of solute per unit time per unit area in a completely stirred solution of large volume would be constant at all times. The remaining processes are (c) and (d). The appropriate equations for the experimental arrangement of these authors follow.

(i) *System (c): Unstirred Layer, Thickness $2l$, Surface at $x = -l$*

$$D \frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t}, \quad -l < x < l, \quad t > 0, \quad \dots \dots \dots (36)$$

$$c = 0, \quad -l < x < l, \quad t = 0, \quad \dots \dots \dots (37)$$

$$c = 0, \quad x = l, \quad t \geq 0, \quad \dots \dots \dots (38)$$

$$c = c_s, \quad x = -l, \quad t \geq 0. \quad \dots \dots \dots (39)$$

The concentration c_e is that concentration of solution which would be in equilibrium with a monolayer of a given surface pressure, that is, $c_e = (\Gamma_0/\delta) \exp(-G/RT)$, where Γ_0 is the surface-excess.

The solution is

$$\frac{c}{c_e} = \frac{1-y}{2} + \sum_{n=1}^{\infty} \frac{(-1)^n e^{-\kappa n^2 \pi^2 t/l^4}}{n} \sin n \frac{\pi}{2} (1-y), \quad \dots \quad (40)$$

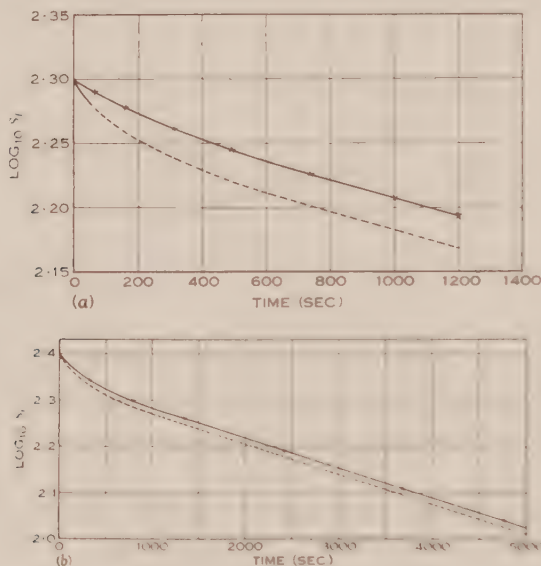


Fig. 3.—The decrease in area (S_f) with time of a film of lauric acid spread on 10^{-3} N HCl. The surface pressure is held constant. Experimental points (\times) from data of Guastalla and Mibashan-Saraga (1949) and Saraga (1950).

(a) Surface pressure 4.0 dyne cm^{-1} . Full curve, $l = 5.73 \times 10^{-2}$ cm., $E = 15.7RT$ (eqn. (50)); broken curve, $l = 6.3 \times 10^{-2}$ cm. (eqn. (41)). Correct slope in steady state only.

(b) Surface pressure 7.1 dyne cm^{-1} . Full curve, $l = 5.76 \times 10^{-2}$ cm., $E = 15.7RT$ (eqn. (50)); broken curve, $l = 6.1 \times 10^{-2}$ cm. (eqn. (41)). Correct slope in steady state only.

where $y = x/l$, $z = D/l^2$. The amount transferred per unit area of surface is

$$A = \int_0^l - \left(D \frac{\partial c}{\partial x} \right)_{x=0} dt, \\ = \frac{Dc_e l}{2} - \frac{4}{\kappa \pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n e^{-\kappa n^2 \pi^2 t/l^4}}{n^2}, \quad \dots \quad (41)$$

If the initial area of the film is S_0 then the area S at any time t is given by

$$\log_{10} S = \log_{10} S_0 - \frac{A}{2.303 \Gamma_0}, \quad \dots \quad (42)$$

where Γ_0 is the constant surface excess. Γ_0 is obtained from the surface pressure, the data of Frumkin (1925), and the curves presented by Rideal and Schofield (1925).

(ii) *System (d): An Energy Barrier, E, to Transfer*

$$D \frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t}, \quad -l < x < l, \quad t > 0, \quad \dots \quad (43)$$

$$c = 0, \quad -l < x < l, \quad t = 0, \quad \dots \quad (44)$$

$$c = 0, \quad x = l, \quad t \geq 0, \quad \dots \quad (45)$$

$$c_s = c_e, \quad x = -l, \quad t \geq 0, \quad \dots \quad (46)$$

$$D \frac{\partial c}{\partial x} + H(c_e - c_s) = 0, \quad x = -l, \quad t > 0, \quad \dots \quad (47)$$

$$-H = Z_f e^{-E/RT}, \quad \dots \quad (48)$$

The solution is

$$\frac{c}{c_e} = \frac{1-y}{2 + \frac{1}{h}} - \sum_{n=1}^{\infty} \frac{-h e^{-\kappa \alpha_n^2 t} \sin \alpha_n (1-y)}{\sin 2\alpha_n \left[\alpha_n^2 + \frac{h(2h+1)}{2} \right]}, \quad \dots \quad (49)$$

where $y = x/l$, $\kappa = D/l^2$, $h = Hl/D$, and α_n are the roots of $\alpha \cot 2\alpha + h = 0$. The amount transferred per unit area is

$$A = l c_e h^2 \left\{ \frac{\kappa t}{h(2h+1)} + \sum_{n=1}^{\infty} \frac{1 - e^{-\kappa \alpha_n^2 t}}{\alpha_n^2 \left[\alpha_n^2 + \frac{h(2h+1)}{2} \right]} \right\}, \quad \dots \quad (50)$$

and equation (42) is used to express the result as a change in area.

(iii) *Numerical Calculations*

In the steady state

$$\frac{d \log_e S}{dt} = - \frac{1}{\Gamma_0} \left(\frac{dA}{dt} \right)_{t \text{ large}}, \quad \dots \quad (51)$$

The values of dA/dt are obtained from equations (41) and (50). From the experimental slope the thickness, $2l$, is deduced. To evaluate equation (50) a second parameter (E) was deduced from a point on the experimental curve. The results are plotted in Figure 3. Thus the system (c) does not fit the results and only system (d) is satisfactory (equation (50)). An activation energy of about $16RT$ is required to fit the results.

IV. PLAUSIBILITY OF THE MODELS USED IN THE SYSTEMS

The stirring of liquid in the experiments of Addison and Hutchinson must arise largely from convection currents. Only small temperature differences are expected in the solution so that the motion will be of the type described by Bénard (1901). Figure 4 illustrates the cellular motion which is developed when a layer of fluid is heated from below or cooled, e.g. by evaporation, from above. When the surface of the fluid (liquid) is free the distribution of velocity in the surface is given by Figure 4 (b). If the upper surface is solid, e.g. liquid held between plates as in the experiments of Saunders and Schmidt (1938), the velocity at the surface is zero. Thus to obtain an unstirred layer it is necessary

to have a surface which is "rigid". Previous experience has shown that monolayers do confer a type of rigidity to the surface. For example, movement of a monolayer over a surface produces a boundary layer (Schulman and Teorell 1938; Crisp 1946). Likewise bubbles of air rising in solutions of surface-active compounds show a terminal velocity characteristic of a solid sphere and not that of a gas sphere (see Lamb 1932; Gorodetzkaja 1949; Levich 1949). Figure 4 (c) shows a probable arrangement in the surface of a liquid layer being stirred by convection: the area above the down-flow sections will have lower

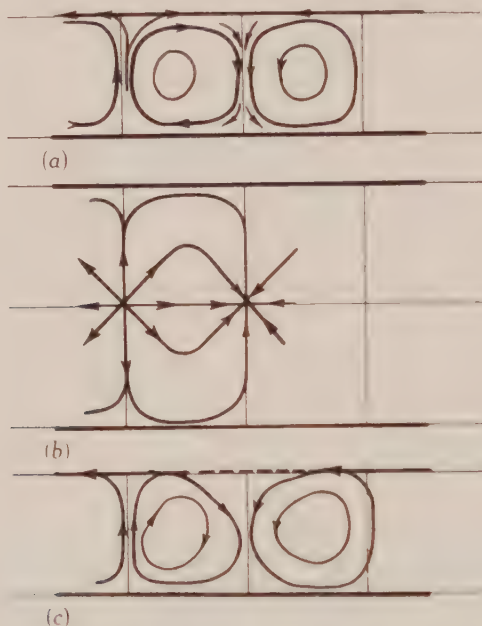


Fig. 4.—Convection patterns in a trough in which the upper surface is cooled.

- (a) Side elevation showing circulation (fixed bottom surface, free upper surface).
- (b) Plan of surface (fixed walls).
- (c) Side elevation showing patterns in the presence of a surface film (fixed bottom surface). The broken portion of the upper surface is quiescent due to the film.

surface tensions than near the up-flow sections because movement across the surface will pack molecules above the down-flow sections. As a consequence the surface is more "rigid" (retarding flow beneath it) near the down-flow sections. It is obvious that the description of such a system by a model of an unstirred layer with well-stirred liquid below is extremely crude. No strict physical interpretation of the thickness of layer or of any activation energy deduced from such data seems possible. This conclusion is supported by the very large difference between the thicknesses of the unstirred layer—from 10^{-3} to 10^{-1} cm.—deduced from the two sets of data examined herein. Much of this difference will arise from the different dimensions of the troughs used in

the experiments as well as differences in rates of evaporation and the heating arrangements in the thermostat.

Rayleigh (1916) and subsequently Jeffreys (1928) examined theoretically the conditions leading to cellular flow. The theory developed by Jeffreys was confirmed by Saunders and Schmidt (1938). Cellular motion commences if the quantity, λ , given by

$$\lambda = \frac{g\alpha s(\theta_1 - \theta_2)h^3}{k\nu}, \quad \dots\dots\dots (52)$$

exceeds 1700, where α is the coefficient of thermal expansion, g the gravitational acceleration, s the specific heat, k the thermal conductivity, ν the kinematic viscosity, h the depth of fluid layer and $\theta_1 - \theta_2$ (>0) the temperature difference between bottom and top. Using the appropriate values at a mean temperature of 20 °C. for water

$$\lambda = 14090(\theta_1 - \theta_2)h^3. \quad \dots\dots\dots (53)$$

Since h is not less than 1 cm. and $\theta_1 - \theta_2$ is probably of the order of 1 °C. in the experiments of the above authors then it is obvious that cellular motion must have been important.

The data of Saunders and Schmidt (1938) show that for a trough 1 cm. deep with $\theta_1 - \theta_2 = 1$ °C. the equivalent system for thermal conduction is an unstirred layer 1/3 cm. thick, the remainder of the liquid being well stirred, that is, it offers no thermal resistance. Now the necessary condition that flow be important is that the flux of heat due to convection shall be greater than or comparable to the flux due to thermal gradient, i.e. $u\rho s \geq k(\partial\theta/\partial x)$, where u is the velocity of flow in the x -direction, s the specific heat, and ρ the density. A similar relation will hold for the solute which is diffusing, i.e. $uc \geq D(\partial c/\partial x)$, where c is the concentration. If u is sufficiently small there is a negligible contribution to the total thermal flux by convection. Since D is some thousand times smaller than k even a small velocity will contribute substantially to the movement of solute. As an approximation we may diminish the thickness of the unstirred layer from 1/3 cm. by the fraction $(D\rho s)/k$, and allowing a parabolic distribution of flow approaching the surface we find an unstirred depth of about 10^{-3} cm. This is reasonable for the data of Addison and Hutchinson (10^{-3} cm.) but not for those of Guastalla and Saraga (10^{-1} cm.). Factors such as temperature difference, concentration gradient, and difference of surface rigidity are, however, of paramount importance and these may have differed widely in the two laboratories.

(a) *Effect of Piston Movement on Movement of Solute*

In the experiments of Guastalla and Saraga the piston maintaining constant pressure on the monolayer of lauric acid causes a downward flow of liquid.*

* This may be observed by spreading over the surface of water a small amount of dye, e.g. crystal violet and some surface-active agent (cationic or non-ionic). The surface appears as a thin blue line when observed horizontally. If the surface is now compressed with a piston the line thickens and the surface behind the piston is left quite free of dye, i.e. there is no observable flow past the piston and solute is moved down into the solution.

Consider only the steady state shown in Figure 2. That is,

$$\frac{d \log S}{dt} = -k, \quad \dots \dots \dots (54)$$

where S is the area at any time t , and k is a constant. If the trough is of width w cm. and the length of film is z cm. then

$$\frac{dz}{dt} = -kz,$$

or

$$v_0 = -kz, \quad \dots \dots \dots (55)$$

where v_0 is the velocity of the piston.

Let δ be the thickness of the boundary layer or depth of trough whichever is the lesser. The velocity at any depth x is

$$\frac{v}{v_0} = \left(1 - \frac{x}{\delta}\right)^2,$$

that is

$$v = -kz \left(1 - \frac{x}{\delta}\right)^2. \quad \dots \dots \dots (56)$$

For a non-compressible liquid we have

$$\frac{\partial v}{\partial z} + \frac{\partial u}{\partial x} = 0, \quad \dots \dots \dots (57)$$

where u is the velocity normal to the surface.

From equations (56) and (57)

$$u = kx \left(1 - \frac{x}{\delta} - \frac{x^2}{3\delta^2}\right). \quad \dots \dots \dots (58)$$

In the steady state

$$D \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial z^2} \right) - \left(u \frac{\partial c}{\partial x} + v \frac{\partial c}{\partial z} \right) = 0. \quad \dots \dots \dots (59)$$

Now

$$\frac{\partial^2 c}{\partial z^2} < \frac{\partial^2 c}{\partial x^2} \quad \text{and} \quad v \frac{\partial c}{\partial z} < u \frac{\partial c}{\partial x},$$

so that equation (59) reduces to

$$D \frac{\partial^2 c}{\partial x^2} - u \frac{\partial c}{\partial x} = 0, \quad \dots \dots \dots (60)$$

with boundary conditions

$$\left. \begin{array}{l} c=0, \quad x=l, \\ c=c_e, \quad x=0. \end{array} \right\} \quad \dots \dots \dots (61)$$

Substituting for u from equation (58) and considering the liquid well stirred at depths greater than l we obtain

$$\frac{c}{c_e} = \int \exp \left\{ \frac{hy^2}{12} (6 - 4y + y^2) + m \right\} dy + q, \quad \dots \dots \dots (62)$$

where $h = (k\delta^2)/D$, $y = x/\delta$, and

$$e^{-m} = - \int_0^{l/\delta} \exp \left\{ \frac{hy^2}{12} (6 - 4y + y^2) \right\} dy,$$

and q is determined from the boundary conditions (61). The amount of solute which leaves the surface per unit area per unit time is

$$A = \left(-D \frac{\partial c}{\partial x} + uc \right)_{x=0},$$

$$= \frac{Dc_0}{\delta \int_0^{l/\delta} \exp \left\{ \frac{hy^2}{12} (6 - 4y + y^2) \right\} dy}. \quad \dots\dots\dots (63)$$

For l/δ small (as it will be) then

$$A = \frac{Dc_0}{l + \frac{k\delta}{6D} \zeta^3 (1 - \frac{1}{2}\zeta)}, \quad \dots\dots\dots (64)$$

where $\zeta = l/\delta$. The value of l required to fit the data is 0.118 cm. This is smaller than that given by equation (41) by some 6 per cent., flow decreasing the solute movement by decreasing the concentration gradient.

The problem with an energy barrier present at the surface may be solved in the same way for the steady state, the amount leaving the surface being given by

$$A = \frac{c_0}{\frac{1}{H} + \frac{1}{D} \left[l + \frac{k\delta}{6D} \zeta^3 (1 - \frac{1}{2}\zeta) \right]}, \quad \dots\dots\dots (65)$$

where H is given by equation (48). The data are insufficient to evaluate both l and H (or E) as we only know the slope of the curve in the steady state.

Thus we may conclude that a system with a stirred layer with or without an energy barrier preventing entry of molecules gives a plausible description of the experimental results of Addison and Hutchinson. It seems very likely that an energy barrier is necessary for the data of Guastalla and Mibashan-Saraga (1949) but as the movement of the piston also decreases desorption this is not certain.

V. DESIGN OF EXPERIMENTS

From the above considerations it is clear that where convective stirring is present the system can only be approximately described by the model of a completely stirred liquid with an unstirred surface layer. There is no adequate description yet of the velocity at every point in a liquid being stirred by convection. Even when such a description is found the equations governing transport are at present insoluble. It is essential therefore to eliminate stirring by reducing temperature inequalities and evaporation to a minimum. More important, however, is the use of troughs of small depth. Equation (52) makes it clear that if the depth, h , of the liquid layer is reduced to say 0.3 cm.

then $(\theta_1 - \theta_2)$ can be as large as 4.5° C. before cellular motion begins. Some difficulty may be experienced in using the Wilhelmy plate under these conditions but this would be a minor inconvenience. For conditions free of convection currents and depth of solution greater than 1 mm. equations (5) or (15) will describe the rate of adsorption.

The type of experiment used by Guastalla and Mibashan-Saraga (constant surface pressure) will require the same trough of small depth if a convective-free system is required. Under these conditions a steady state will *not* be attained and consequently equations (64) and (65) cannot be applied. It would seem wiser, therefore, to conduct the experiments at constant surface area unless a mathematical solution be obtained for the non-steady conditions with a moving piston.

VI. ACKNOWLEDGMENTS

I am grateful to Dr. J. Bowler Reed and to Mr. J. Barker for discussion on points presented in this paper.

VII. REFERENCES

- ADDISON, C. C., and HUTCHINSON, S. K. (1949).—*J. Chem. Soc.* **1949**: 3387.
BÉNARD, H. (1901).—*Ann. Chim. Phys.* **8**: 62.
CRISP, D. J. (1946).—*Trans. Faraday Soc.* **42**: 619.
FRUMKIN, A. N. (1925).—*Z. Phys. Chem.* **116**: 466.
GORODETZKAJA, A. (1949).—*Zh. Fiz. Khim.* **23**: 72.
GUASTALLA, J., and MIBASHAN-SARAGA, L. (1949).—"Surface Chemistry." p. 103. (Butterworth & Co. Ltd.: London.)
JEFFREYS, H. (1928).—*Proc. Roy. Soc. A* **118**: 195.
LAMB, H. (1932).—"Hydrodynamics." (Cambridge Univ. Press.)
LEVICH, B. G. (1949).—*Zh. eksp. teor. Fiz.* **19**: 18.
RAYLEIGH, LORD (1916).—*Phil. Mag.* **32**: 529.
RIDEAL, E. K., and SCHOFIELD, R. K. (1925).—*Proc. Roy. Soc. A* **109**: 57.
RIDEAL, E. K., and SUTHERLAND, K. L. (1952).—*Trans. Faraday Soc.* (in press).
SARAGA, L. (1950).—*C.R. Acad. Sci. Paris* **231**: 46.
SAUNDERS, O. A., and SCHMIDT, P. J. (1938).—*Proc. Roy. Soc. A* **165**: 216.
SCHULMAN, J. H., and TEORELL, J. (1938).—*Trans. Faraday Soc.* **34**: 1337.
WARD, A. F. H., and TORDAI, L. (1946).—*J. Chem. Phys.* **14**: 453.

THE RÔLE OF SURFACE TRANSPORT IN THE STABILITY AND BREAKDOWN OF FOAMS

By W. E. EWERS* and K. L. SUTHERLAND*

[*Manuscript received July 3, 1952*]

Summary

A new theory of foam stability is proposed which demonstrates that the transport of substrate, accompanying a movement of the surface of the bubble film, is a dominant factor in the stability of foams and in the action of foam breakers. The surface moves from a region of low surface tension (high surface pressure) to a region of high surface tension. The surface tension gradients arise from disturbances which may be caused by mechanical or thermal shocks, or by the addition to the surface of particles, droplets, or vapour of a surface-active material. When the surface tension is highest at the centre of disturbance the film will be stable; when the surface tension is lowest at this point the surface film and hence the substrate will be moved away from this point and the film will rupture.

I. INTRODUCTION

It has been generally recognized for many years that the thin liquid films which comprise a foam are thermodynamically unstable. However small the free energy of the surface of the film may be, there is a decrease in the total free energy of the system when a foam breaks up into a few droplets of very much smaller total surface area. The extraordinary metastability of some liquid films, which enables them to survive under carefully controlled conditions for periods up to 3 years, has for long excited the interest of physical chemists. Thus Gibbs made a most important theoretical contribution as long ago as 1878. Nevertheless in the breakdown of films and, in particular, in the action of froth breakers there remain many unexplained phenomena, and the subject as a whole lacks a generally satisfactory theory.

In the present paper we will introduce to this subject the concept of surface transport, the movement of substrate which accompanies the movement of a surface film. This phenomenon, recognized by van der Hönert (1932), Harkins and Kirkwood (1938), Schulman and Teorell (1938), and Crisp (1946), has not been considered hitherto as a dominant factor in foam stability and foam breakdown. We propose to review the significant literature on foams and foam breakers, then to propound the new theory and to examine the available data from this point of view. It will be shown that the new hypothesis, taken together with the statement by Gibbs of the "equilibrium" conditions for stability, and giving due weight to the importance, in particular systems, of such factors as viscosity, surface viscosity and plasticity, and evaporation, provides a consistent explanation of the mechanical properties of foams.

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

II. REVIEW OF CURRENT THEORIES

(a) Theories of Foam Stability

Easily the most important contribution to the present knowledge of why foam films may persist for long periods is that of Gibbs (1878). Although he did not specifically discuss an unusual effect of evaporation in certain systems or the effects of surface viscosity, air permeability, and electrostatic repulsions between charged surfaces, Gibbs foreshadowed all the theories which have been proposed and gave the clearest statement of most of them. Because the value of his contribution has not been generally recognized and, in fact, has been ignored or misunderstood by many, including the authors of some text-books, we will describe the Gibbs theory in some detail.

Gibbs considered film stability as an "equilibrium" condition. He showed that, in order to support its own weight, a film must be capable of maintaining small differences in surface tension over its surface. In a vertical film it can be readily calculated that for a thickness of δ cm. the difference $\Delta\gamma$ in surface tension at two points, one of which is h cm. above the other, is as given in Table I.

TABLE I
THE VARIATION IN SURFACE TENSION ($\Delta\gamma$ DYNE CM.⁻¹) REQUIRED TO SUPPORT
A FILM OF THICKNESS δ AND HEIGHT h

δ (cm.) \ h (cm.)	1	5	50
	(Δγ dyne cm. ⁻¹)		
10 ⁻³	0.5	2.5	25.0
10 ⁻⁴	0.05	0.25	2.5
10 ⁻⁵	0.005	0.025	0.25
10 ⁻⁶	0.0005	0.0025	0.025

It will be noted that the magnitude of the surface tension difference decreases directly both with the thickness of the film and with its height. It is significant that the most persistent films are horizontal and that they thin to the "black" so that the surface tension differences required to support their mass is very small.

Gibbs did not explain why the surfaces of a vertical film do not flow upward under the influence of the surface tension gradient. We can show, however, that this movement is prevented by the viscous drag of the liquid draining down the centre of the film. Indeed this bulk flow can be regarded as a necessary condition for maintaining the surface tension gradient. This may be expressed as the boundary condition at the surface ($y=0$)

$$\partial\gamma/\partial x + \eta\partial u/\partial y = 0,$$

to an equation describing flow in the film. Here η is the viscosity, u the velocity of an element in the direction x , vertically downwards, and y is the distance into the film from its surface.

Gibbs defined a quantity E , the elasticity of the film, by the equation

$$E = 2A(\partial\gamma/\partial A)_{\text{const. } G_1, G_2},$$

where G_1 and G_2 are the total quantities of the components 1 and 2 in the film and A the area of the film. The elasticity is a measure of the increase of surface tension in response to an increase in area. As the film is stretched the surface concentration falls and the solute content of the bulk of the film will be decreased in partially restoring the surface concentration. Gibbs made it clear in his discussion, that E depends on the equilibrium surface tension of the stretched film and *not* the non-equilibrium surface tension. He noted that the restoring force may be greater by virtue of the non-equilibrium surface tension of the newly stretched film, but did not stress this point. (In this respect he has been misquoted by a number of authors.)

From the foregoing it will be recognized that the bubble films are formed only where the liquid contains more than one component other than the gas.

Gibbs recognized that the drainage of a bubble film is controlled by viscous flow under pressure gradients caused by gravity and by the high curvature of the so-called Gibbs ring around the periphery of the film. He drew attention also to the portion of the film adjacent to the Gibbs ring. In this region the liquid is drawn faster into the Gibbs ring than it can be supplied by drainage from other parts of the film. Thus very thin patches occur which, when situated below thicker portions of the film, are unstable and rise in the film until they reach a stable position, that is, in a portion of the film of the same thickness. Thus the suction of the Gibbs ring is an important factor in the thinning of the whole film.

The effects of evaporation and condensation of components of the film were also discussed by Gibbs and he concluded that where evaporation of the solvent is possible the diminution in the thickness of the film will be accelerated; and if the surface-active component evaporates and recondenses at the top of the film, breakdown will be very rapid.

Based as it is on the classical thermodynamics of the equilibrium state, Gibbs's treatment suffers from the defect that, in films, equilibrium is rarely achieved or even approached. The treatment does not, therefore, give due emphasis to non-equilibrium properties such as, for example, the dynamic surface tension. Furthermore it lacks, albeit inevitably, a molecular picture of how the liquid is returned to a locally thinned portion of the film to restore uniform thickness. The first defect has been partially remedied by the insistence of Edser (1922) on the importance of the non-equilibrium surface tension. In the present paper we will endeavour to explain the mechanism by which a local distortion is healed.

Edser (1922) emphasized that when a bubble film is distorted, the portion of the surface which is stretched is not in equilibrium with the underlying liquid; its surface tension is higher than the equilibrium value, and the restoring force is greater than the elasticity defined by Gibbs. The validity of Gibbs "equilibrium" condition for film stability is unaffected, but Edser described more completely the response of a film to sudden local stretching. Other

authors, including Foulk (1931), have also stressed the importance of the dynamic surface tension but, because no satisfactory method is known for the measurement of dynamic surface tension, quantitative confirmation is lacking. Edser (*loc. cit.*) and Bureik (1950) have suggested that the *rate* of attainment of surface tension equilibrium is also important. Wilson and Ries (1923) and Talmud and Suchowoloskaja (1931) suggested that either high rigidity, high plasticity, or high viscosity in the surface layer is necessary for film stability. Brown, Thuman, and McBain (1949*a*) have reported fair correlation between stability and surface viscosity, but drew the conclusion that high surface viscosity alone is not sufficient to produce a stable foam.

Neville and Hazellhurst (1937) observed the froths on a number of "pure" organic liquids. Although their results are at variance with all reliable work on pure liquids, and the purity of their materials must be considered suspect, they noted that, in these systems, the film lives were increased when free evaporation was allowed and decreased when the films were exposed to vapour of the film-forming liquid. They proposed the ingenious explanation that when free evaporation is allowed the thinner parts of the film cool most and the higher surface tension resulting causes liquid to move towards these thinner portions. Neville and Hazellhurst considered that a similar evaporation-temperature-surface tension cycle would be relatively unimportant in foams of solutions of surface-active materials, and indeed this must be so since one of the essential conditions for preserving soap films over long periods is that evaporation should be eliminated.

A further suggestion by Brown, Thuman, and McBain (*loc. cit.*) was the duplex theory which stated that two or more solute components were required for very stable foams. While there is good experimental evidence for this contention the mechanism whereby the duplex film causes a very stable froth is not made clear. The same authors (Brown, Thuman, and McBain 1949*a*) demonstrated that stable films were relatively impermeable to air.

McBain *et al.* (1944) suggest that the electrostatic repulsion of the two equally charged sides of the film will resist the collapse of the film.

(b) Theories of the Action of Foam Breakers

At the outset a clear distinction must be drawn between what we shall term "foam inhibitors" and "foam breakers". These terms have often been used synonymously in the past. We will confine the term foam inhibitor to those agents which, by acting as nuclei for the formation of bubbles in the bulk of a liquid, give fewer and larger bubbles. It is the small bubbles which produce the more troublesome foam. The polyamides described by Gunderson and Denman (1948) are foam inhibitors by this definition. Our discussion will deal with "foam breakers", or agents which cause the breakdown of the liquid films which form the foam. Theories of the action of foam breakers have been reviewed recently by Pattle (1950) and by McBain *et al.* (1944). They may be classified in three main groups.

(i) *Displacement*.—Ohl (as cited by Pattle) and Pattle (1950) proposed that a foam breaker, which is itself incapable of producing a stable foam, displaces

the foaming reagent from the surface of the film. However, McBain *et al.* (loc. cit.) quote examples for oil films in which the foam breaker when dissolved in the oil actually produces a stable foam. Another example in which this theory does not hold is quoted on page 709.

(ii) *Spreading*.—Robinson and Woods (1948) correlated a rupture coefficient with a spreading coefficient defined by

$$S = \gamma_A - \gamma_{AB} - \gamma_B,$$

where γ_A and γ_B signify the surface tensions of the froth solution and the froth breaker respectively and γ_{AB} their interfacial tension. Subsequently Pattle (1950) demonstrated a much more convincing correlation between the results of foam breaking tests and spreading tests. He concluded that spreading assists foam breaking merely by extending the distribution of the foam breaking reagent, but states that probably there is no causal relationship between spreading and foam breaking.

Another theory which it is convenient to discuss here is that proposed by Robinson and quoted by McBain *et al.* (1944). Robinson suggests that the polar groups in the foam breaker repel the non-polar oil in oil films. A more credible explanation is that the foam breaker spreads over the oil and, as will be shown later, the substrate is transported away with the spreading molecules until rupture occurs.

(iii) *Inhomogeneity*.—Hughes (as cited by McBain *et al.*) suggested that the action of immiscible liquids as foam breakers may be due to small droplets of oil puncturing the film and interrupting the layer of surface-active material which stabilizes it. McBain cites examples where lenses of a liquid penetrate the film without causing breakdown. This theory does not attempt to account for the action of soluble foam breakers.

McBain *et al.* (1944) conclude that inhomogeneity arising from concentration differences, from the volatility of one or more components and from physical shocks, mechanical and thermal, may cause film breakdown. It will be shown later that all of these circumstances favour surface transport.

III. THE CONCEPT OF SURFACE TRANSPORT

In 1938, Schulman and Teorell demonstrated that when a monomolecular film of oleic acid moves over water a quantity of water moves with it, corresponding to a layer of quite considerable thickness. Thus a monolayer moving at 5.0 cm. sec. along the surface of a reservoir 2 cm. wide, and containing initially 10 ml. water, emptied the reservoir completely in 6.5 minutes. The same effect was recognized implicitly by Harkins and Kirkwood (1938) when, in deriving their formula for the measurement of surface viscosity by the canal method, they assumed that there was no slip between the moving film and the substrate immediately in contact with it. This assumption, that there shall be no discontinuity in the velocity gradient at the boundary, is in conformity with all evidence from other branches of hydrodynamics.

Another demonstration of surface transport, which was not recognized as such at the time, was given by McBain *et al.* (1944). A drop or crystal of a foam

breaker was placed on a film of oil spread on glass. The oil moved away from the drop or crystal leaving the glass surface surrounding it apparently bare. Similar experiments have been performed by Evans (unpublished data) in this Laboratory. A crystal of sodium hexadecyl sulphate was dropped onto a wet glass slide. Spreading of the sodium hexadecyl sulphate was observed by watching the motion of talc on the surface of the film, and as in McBain's experiment the liquid layer was completely removed from an area surrounding the crystal.

Surface transport was studied also by Crisp (1946). He investigated the dependence of surface flow on the dimensions of the channel, the surface pressure gradient, the viscosity of the substrate, and the character of the film. A number of films were examined, but in no instance was there any evidence of slip between the film and the underlying liquid.

We will consider the flow in a bubble film where a uniform surface pressure gradient is applied only in one side of the film. It is reasonable to assume, as a first approximation, that the other side will be stationary because any flow in it will be opposed by an induced surface pressure gradient. In these circumstances the flow will be that of a liquid subjected to shear between parallel planes. The velocity of an element of liquid between the planes will vary linearly with the distance from one of them. It would follow, therefore, that the average thickness of a film, which might be considered to flow at the same velocity as the moving surface, is equal to half the total thickness of the film. The validity of this conclusion as applied to bubble films could be affected by three assumptions which we have made. Firstly, the surface pressure gradient in bubble films will not usually be uniform but may arise, for example, from a point source of surface-active material. Secondly, it has been assumed implicitly that the system has reached a condition of steady flow. With bubble films we will often be concerned with the initial movements in which the inertia of the film would be involved. This would tend to reduce the amount of flow. Finally, the assumption that the other surface of the film is immobile may be suspect. However, any movement of this surface will permit a greater quantity of the bulk of the film to flow with the upper surface.

To summarize, it has been established that the movement of a surface film causes a considerable movement of the substrate. This conclusion is reached entirely from hydrodynamic considerations assuming no anomalous viscosity in the surface layer: no long range orientation in the water within the film need be invoked to account for the flow.

IV. THEORY OF FOAM STABILITY

It must be a central postulate in any theory of foams that the bubble films break by thinning. Only when the thickness of a portion of the film is reduced to 50 Å, or less, does it appear feasible for random motions of individual molecules, particularly molecules of the surface film, to cause sudden breakdown of the film.

Factors which are known to affect the rate of thinning, and hence the ability of the film to withstand shocks, are listed below.

- (i) Viscosity (and/or plasticity) in the surface or the bulk of the film. High viscosity increases the persistence of films by reducing the thinning rates and by dissipating shocks.
- (ii) Volatility of components of the film. In so far as low volatility conserves the materials which constitute the film its effect is well recognized. Other effects of evaporation will be discussed later.
- (iii) Permeability of the film to gas. Here again the conservation of the material components of the foam is involved, for if the gas escapes through the bubble walls the froth will collapse.
- (iv) Variation of the surface tension on stretching. This factor is recognized to be important but no detailed mechanism has been described of how a film, thinned by stretching, may restore itself.
- (v) Electrostatic repulsion between similarly charged film surfaces. It will be conceded that when the thickness of a film is such that there is interference between the electrical double layers associated with the surfaces, that is, when the film thickness is less than $c. 1000 \text{ \AA}$ (depending on the ionic strength of the solution), further thinning will be opposed by the electrostatic repulsion between the surfaces. This effect will be reduced by any factor which tends to decrease the thickness of the double layer. Thus high electrolyte concentration would tend to reduce foam stability, a conclusion which is consistent with the experimental results of Bartsch (1924) who worked with alcohol frothers.

It is evident that of these factors only (iv) and perhaps (v) can operate in restoring a film which is thinned locally almost to the point of rupture. We will consider mainly the stability of films in systems wherein factors (i), (ii), (iii), and (v) are fixed, though in some instances the secondary effect of variations in one or more of these factors on surface transport will be noted.

The theory of the rôle of surface transport in foam stability may be stated as follows:

Whenever a surface tension gradient, $\partial\gamma/\partial x$, is set up in the surface of a bubble film the surface will tend to move from the region of lower surface tension to the region of higher surface tension, that is, the surface will move up the surface tension gradient. *Any such motion will be accompanied by the movement of a considerable part of the liquid comprising the film.* If the disturbance or deformation which initiates the surface pressure gradient is such that $\partial\gamma/\partial x$ is negative, x being measured from the centre of the disturbance, the film will be stable. If $\partial\gamma/\partial x$ is positive the film will be unstable.

Surface tension gradients may be set up by mechanical deformations (for example, stretching) in which a change of area is involved, by thermal shocks, or by the addition to the film from an external source of surface-active material which may or may not be identical with the foam stabilizing material. Disturbances of the first two kinds, which do not involve the addition of material

to the film, will be discussed in relation to foam stability. The third type of disturbance is typical of the action of foam breakers.

(a) *Mechanical Deformation—Stretching*

An obvious method by which a film may be broken is to extend it until its thickness is about 50 Å, that is, so small that spontaneous rupture may occur by the random movement of its component molecules. It is instructive to consider how much extension of the film is necessary to reduce its thickness to 50 Å. Thus a film 10^{-4} cm. thick must be stretched to 200 times its area, that is, to 14 times its length in two directions at right angles. An extension of this magnitude is unlikely even when the stretching is localized to a small area of the film, and it appears probable that mechanical deformation is an important mechanism of film breakdown only when the film is already very thin. We have already referred to Gibbs's description of the irregular manner in which films are thinned. The thinner regions near the top of a vertical tube or adjacent to the Gibbs ring will be more sensitive to breakdown by localized mechanical shock than other regions.

The effect of stretching a portion of the film is primarily to deplete the number of adsorbed molecules per unit area of surface. This will cause a local increase in surface tension, which will cause in turn surface flow towards the stretched film. The flow of substrate accompanying the flow of the surface layer will tend to return the stretched portion of the film toward its former thickness. The effect of surface transport in this system is thus to prevent local thinning and to ensure that the normal viscous thinning proceeds uniformly.

There are two factors in this process which must be considered: the rate of adsorption of the surface-active reagent and the rate of spreading. If the adsorption is rapid, the deficiency of surface-active molecules in the stretched surface may be made up largely by adsorption from the bulk of the film. This will destroy the surface tension gradient and the amount of solution returned by surface transport will not be sufficient to restore the film. Their comparatively high rate of achieving surface tension equilibrium (Addison 1945; Rideal and Sutherland 1952) would account for the low stability of froths of alcohol solutions. In this respect the alcohols may be contrasted with the paraffin-chain salts, which in concentrations up to the critical concentration for micelle formation have much lower rates of adsorption (Nutting, Harkins, and Long 1940), but which above this concentration are adsorbed more rapidly and show a corresponding decline in froth stability (Bureik 1950). A study of systems in which the rate of adsorption is changed by the addition of some other component would test this explanation, though one would need to avoid incidental changes of other factors (for example, the thickness of the electrical double layer).

For soluble films the rate of spreading is governed *inter alia* by the viscosity of the substrate, the surface viscosity of the adsorbed film and, of course, by the rate of adsorption insofar as it controls the persistence of the surface tension gradient. The adsorption rate also may be controlled by the bulk viscosity which limits the rate of diffusion of the adsorbate into the surface region.

The rôle of high viscosity, either in the surface or the bulk, is usually to dissipate mechanical shocks by opposing both the disruptive and the restoring forces. As high bulk viscosity will reduce both the rates of distortion and adsorption it is invariably a stabilizing factor, whatever the magnitude of the surface viscosity. High surface viscosity, on the other hand, when associated with low bulk viscosity may lead to more complex behaviour. In systems in which the surface tension gradients, produced by local stretching, are large enough to cause surface transport, high surface viscosity will reduce the rate of spreading although the adsorption of the frother from the bulk is not impeded. Thus the surface tension gradient may disappear before the film can be healed by surface transport. This leads to the conclusion that high surface viscosity alone is insufficient for foam stability, a conclusion which is consistent with the experiments reported by Brown, Thuman, and McBain (1949*a*) using dodecanol. Another type of system is that in which the surface shows plasticity and the bulk viscosity is relatively low, for example, films of saponin solution. Here the surface tension gradients which would be produced by local stretching are negligible, so that surface transport is a negligible factor in the stability of the film. The film owes its stability against mechanical shock almost entirely to its plasticity. A local deformation will be removed very slowly as the surface tension reduces the total area of the film to a minimum consistent with its boundaries. A good example is shown by Boys (1924); the deflation of a saponin bubble leads first to a crinkled cone which changes slowly to a hemisphere.

The rate of stretching may be quite important. If we imagine an element of surface to be stretched to a given *extent* at different rates then the greater the rate, the greater will be the surface tension gradient because there is a shorter time for the solute to be adsorbed, and the greater the resilience of the film to mechanical shock. If, however, the element is stretched for a given *time* at different rates, no method is apparent by which the future of the element may be predicted.

(b) *Thermal Deformation—Constant Area*

If a small portion of the film becomes hotter or colder than the surroundings, a surface tension gradient, $\partial\gamma/\partial x$, will be set up according to the sign of the gradient, $\partial\gamma/\partial T$. If the temperature change is accompanied by, or caused by evaporation, the change of surface tension with the activity of the solute, $\partial\gamma/\partial a$, must also be taken into account. In Table 2 the stability of the film is given, having been determined from the sign of $\partial\gamma/\partial x$ for all possible values of ΔT , the temperature change, Δa , the associated change in the activity of the solute, $\partial\gamma/\partial T$, and $\partial\gamma/\partial a$. Cooling as a result of the evaporation of solute has been ignored as trivial. The effects of condensation are not listed as these are the exact reverse of the effects of evaporation, the signs of ΔT and Δa and hence of $\partial\gamma/\partial x$ being reversed. It is assumed throughout that $\partial\gamma/\partial a$ is negative.

The interesting features of Table 2 are as follows :

- (i) For the common condition that surface tension decreases with increasing temperature, a local increase in temperature will lead to instability.

- (ii) Usually evaporation of a surface-active solute accompanied by an increase in temperature favours stability.
- (iii) Evaporation of the solvent, on the other hand, whether accompanied by heating or cooling, appears to be a fruitful source of film breakdown, not merely because of the dissipation of material, but because the surface tension gradients produced cause liquid to be transported away from the region where the solvent is evaporating.

TABLE 2
STABILITY OF FILM UNDER THERMAL DEFORMATION

Main Condition	Subsidiary Condition	With Evaporation			
		Without Evaporation $\Delta T > 0$	Solute* Evaporating $\Delta T > 0$	Solvent Evaporating	
				Evaporation causes Cooling $\Delta T < 0$	Heating Effect > Evaporative Cooling $\Delta T > 0$
$\partial\gamma/\partial T > 0$	$\partial\gamma/\partial T \Delta T > \partial\gamma/\partial a \Delta a$	Stable	—	—	—
	$\partial\gamma/\partial T \Delta T < \partial\gamma/\partial a \Delta a$	—	Stable	Unstable	Stable
$\partial\gamma/\partial T < 0$	$\partial\gamma/\partial T \Delta T > \partial\gamma/\partial a \Delta a$	Unstable†	—	—	—
	$\partial\gamma/\partial T \Delta T < \partial\gamma/\partial a \Delta a$	—	Unstable	Stable†	Unstable
		—	Stable	Unstable	Unstable

* Reference has already been made to the case considered by Gibbs of isothermal evaporation of the solute and recondensation on the upper part of the film. This action would tend to remove the surface tension gradient from top to bottom of a vertical film, so that the Gibbs equilibrium condition for a stable film would not be satisfied. This is not a surface transport phenomenon, because the difference in vapour pressure arises initially from the surface tension gradient, a gradient which could not be reversed, but merely reduced by the evaporation-condensation cycle.

† In the system described by Neville and Hazellhurst (1937) if it be assumed that their liquids were not pure but contained relatively high concentrations of similar compounds then $\partial\gamma/\partial T \Delta T = \partial\gamma/\partial a \Delta a$ for the small changes of a accompanying the evaporation and a local lowering of temperature will not produce instability. For a pure liquid the criterion of stability is given by $\partial\gamma/\partial T < 0$, for ΔT negative.

Surface tension gradients could arise also by applying an electrical potential along the surface of the film, but neither this possibility nor the effect of electrokinetic potentials set up by flow in the liquid films has been considered.

V. THEORY OF ACTION OF FROTH BREAKERS

We have discussed already the response of a film under various conditions to mechanical and thermal shocks. It is resilience to such shocks and ability

to regain uniform film thickness which cause foam stability. By contrast we will deal in this Section with the action of reagents which cause the bubble film to rupture.

We will examine the application to foam breakers of the postulate given earlier, that when $\partial\gamma/\partial x$ (the gradient of surface tension with distance, x , measured away from the centre of disturbance) is positive, and remains positive, the film will inevitably break.

Froth breakers will be classified, for the purpose of this discussion, into three groups.

- (i) Solids or liquids containing a surface-active material other than the substance stabilizing the film.
- (ii) Liquids which contain the foam stabilizer in higher concentration than it is present in the foam.
- (iii) Vapours of surface-active liquids.

We do not include reagents which break foams by chemical reaction with the foam stabilizer, for example, mineral acids on soap films.

Group (i).—It will be clear that the surface tension gradient produced by froth breakers of this type can be considered as a spreading pressure. We have already quoted examples to indicate that spreading is accompanied by movement of the substrate.

Let us distinguish two cases: the liquid either spreads as a monolayer or as a lens. In either case it will be assumed that the advancing edge of the spreading film will sweep the film stabilizing the foam before it, so that the spreading film is initially in contact with a newly formed surface of the same composition as the bulk.

The condition that the spreader B will spread initially over the solution A is given by

$$S_{mi} = \gamma_A - \gamma_{B(S)} > 0,$$

for the monolayer spreading, and

$$S_{li} = \gamma_A - \gamma_{SB} - \gamma_B > 0,$$

for spreading as a lens. In these equations S_{mi} and S_{li} are respectively the initial spreading coefficients for monolayer and lens, γ_A is the surface tension of the solution, $\gamma_{B(S)}$ the *initial* surface tension of the monolayer on the *swept* solution, γ_{SB} the *initial* interfacial tension between spreader and *swept* solution, and γ_B the surface tension of the spreader as a liquid. Thus one criterion of a good foam breaker is that S_{mi} or S_{li} should be as large as possible.

For a given element of interface the tensions $\gamma_{B(S)}$ and γ_{SB} may decrease with time if a mixed film is formed between the spreader and ions or molecules of the foam stabilizer which reach the interface from the bulk of the solution. If $\gamma_{B(S)}$ and γ_{SB} decrease it follows that there will be an interfacial tension gradient, between a newly formed portion of the interface and an element formed earlier. This gradient is in the reverse direction from the gradient which initiated spreading, and any movement along this gradient will return substrate to the

portion of the bubble film originally thinned. The condition for reverse spreading between two elements of interface are summarized in the equations:

$$S_{mR} = \gamma_{B(S')} - \gamma_{B(S)} < 0,$$

$$S_{LR} = \gamma_{S'B} - \gamma_{SB} < 0,$$

where $\gamma_{B(S')}$ and $\gamma_{S'B}$ are the tensions of elements of interface which having been formed earliest have become partly equilibrated with the underlying solution. It will follow that for the most effective foam breakers S_{mR} and S_{LR} are close to zero. They will be zero only when no mixed films form.

We must consider also the portion of the adsorbed layer of foam stabilizer which has been compressed by the spreading film or lens. For those foam stabilizers, for example, alcohols which attain surface equilibrium quickly, that is, where any excess above the equilibrium dissolves "immediately" as the film is compressed, there is little or no disturbance of the surface beyond the line which the advancing monolayer has reached. In these circumstances movement of a tale particle on the surface is a good indication of spreading. With other foam stabilizers, notably paraffin-chain salts or proteins, excess molecules dissolve very slowly so that the advance of the spreading film will be retarded by an increasing surface pressure due to the compressed frother film, that is, γ_{LI} is not constant, and even when no mixed film is formed the initial spreading coefficients S_{m_i} and S_{L_i} may decrease with time at a rate dependent on the total area of the film. Thus the effectiveness of a froth breaker may depend upon the area of laminae composing the foam, being less effective when the area is small. Furthermore portions of the bubble film which the spreader has not reached must be moved considerably and the movement of tale on these portions is not a reliable guide to spreading. It is certain that the "rapid spreading" of stearic acid on a film of sodium oleate solution, which was noted by Pattle (1950) was an artefact arising from this cause, so that the failure of stearic acid to break a film stabilized by sodium oleate is consistent with our theory.

It is interesting to discuss the sensitivity to froth breakers of foams of materials such as saponin and various proteins. Here the rate of spreading would be retarded by viscosity of the frother film but the ability of the frother to build up a force resisting spreading is limited by its low collapse pressure. Although the spreading is slow it will proceed inexorably, removing liquid from the bubble wall. There is no mechanism for rapid return of liquid to the thinned film and consequently it breaks.

If the foam breaker is volatile and the atmosphere is not saturated with its vapour, the surface pressure gradient from the drop of foam breaker to the surrounding film must be steepened by evaporation so that breakdown will be accelerated.

Group (ii).—It will be obvious from our discussion that, if differences appear in the concentration of the frother at various parts of the bubble film, surface tension gradients will be set up which will lead to transport of the substrate. This condition frequently arises in a froth column. The uppermost bubbles in the foam will be fairly completely drained. When they burst the concentration

of frother in the droplets produced will be considerably greater than in the films below, which are sprayed by these droplets. Suppose that a bubble film has a surface concentration of 4×10^{-10} mol. cm.⁻² (40 \AA^2 per molecule) and a bulk concentration (excluding the surface) of 2×10^{-6} mol. cm.⁻³. If the film is 10^{-4} cm. thick the total amount of the solute per cm.² of film is 8×10^{-10} moles in the surface plus 2×10^{-10} moles from the bulk. If the surface of the droplets is ignored the concentration would thus be five times greater than the original bulk concentration and this often exceeds the solubility of the frother. It is not surprising, therefore, that the remnants of the burst bubbles will act as froth breakers on the bubbles below. This action once started continues until the entire column of froth is destroyed, a behaviour which is typical of alcohol frothers. The well-known maximum in the curve relating foam stability to concentration of frother of this type we believe is due to this ability of the frother to act also as a froth breaker in the manner described above.

In solutions of pure detergents a similar phenomenon can be observed. Solutions of sodium dodecyl sulphate containing 0.6 g. per litre or less give froths which decay in the typical manner described above. At higher concentrations, however, this mode of breakdown is not seen. At low concentrations of sodium lauryl sulphate the gradient of surface tension with concentration $\partial\gamma/\partial c$ is large and negative, so that a solution of higher concentration has a lower surface tension than, and will spread on, a solution of lower concentration. With detergents, however, as the critical concentration for micelle formation is approached $\partial\gamma/\partial c$ approaches zero. Consequently although the surfaces of the lower bubbles may be sprayed with droplets of much higher concentration of solute, the important and necessary condition, that the surface shall be inhomogeneous in surface tension, is not satisfied.

Group (iii).—The oft-quoted example of ether as a froth breaker falls into the third group, in which we consider vapours of surface-active liquids. There are two cases to be considered; the vapour may condense as a drop on the surface in which a spreading process, as discussed in group (i), is involved; or the vapour, coming in contact with a portion of the film may be adsorbed in greater surface concentration at one point than another, thus giving rise to a concentration gradient similar to those discussed in (ii). It is significant that, as McBain *et al.* (1944) have reported, if the solution or the air is saturated with ether its foam-breaking action disappears. Inhomogeneity of surface tension, leading to surface transport, is a primary necessity for foam breakdown.

VI. CONCLUSION

It has been proposed in the foregoing that the transport of liquid substrate, which accompanies the movement in a surface tension gradient of a portion of the surface of a bubble film, is a dominant factor in both the stability and breakdown of the film. Depending on the sign of the surface tension gradient from the centre of a disturbance to the surrounding film, liquid will move towards or away from the centre and the film will remain stable or it will rupture. To measure the quantities which must be known in order to test the theory adequately presents some real difficulties. Thus we must measure the gradients

which are set up in response to given disturbances. Where the surface is stretched locally the non-equilibrium surface tension should be measured as a function of time, during and after the deformation. All known experimental methods, such as the oscillating jet, will yield results characteristic of the method and the apparatus (Rideal and Sutherland 1952) and not entirely characteristic of the solution. This difficulty arises because transport of the solute to the surface is caused both by diffusion and hydrodynamic flow (Sutherland 1951). Most of the quantities used in Table 2 to describe the response of the film to thermal shock can be measured, with the probable exception of Δa , the change in activity of the solute resulting from the evaporation of either component. In testing the theory of the action of foam breakers real spreading must be distinguished from pseudo-spreading (see p. 708). The more fundamental properties of these systems, for example, the non-equilibrium interfacial tensions measured as a function of time, present difficulties parallel to those referred to above for the air-solution interface. It appears, therefore, that many of the quantities required to test the theory are, at present, inaccessible, and the usefulness of the theory is thereby impaired.

On the other hand the theory provides a consistent qualitative mechanism for the healing and rupture of bubble films and is consistent with the experimental data of the literature.

VII. REFERENCES

- ADDISON, C. C. (1945).—*J. Chem. Soc.* **1945**: 98.
 BARTSCH, O. (1924).—*Kolloidchem. Beih.* **20**: 1.
 BOYS, C. V. (1924).—"Soap Bubbles." (Society for Promoting Christian Knowledge: London.)
 BROWN, A. G., THUMAN, W. C., and MCBAIN, J. W. (1949a).—Surface viscosity of detergent solutions as a factor in foam stability. Tech. Rep. No. 4 under contract N7-onr-321-TO-II between Stanford Research Institute and Office of Naval Research.
 BROWN, A. G., THUMAN, W. C., and MCBAIN, J. W. (1949b).—Transfer of air through adsorbed films as a factor in foam stability. Tech. Rep. No. 5 (ibid).
 BURCIK, E. J. (1950).—*J. Colloid. Sci.* **5**: 421.
 CRISP, D. J. (1946).—*Trans. Faraday Soc.* **42**: 619.
 EDGER, E. (1922).—Brit. Ass. Adv. Sci. 4th Rep. on Colloid Chemistry. Vol. 4. p. 263.
 FOULK, C. W., and MILLER, J. N. (1931).—*Industr. Engng. Chem.* **23**: 1283.
 GIBBS, J. W. (1878).—See "Collected Works of J. Willard Gibbs." pp. 300-14. (Longmans, Green & Co.: New York, 1931.)
 GUNDERSON, L. O., and DENMAN, W. L. (1948).—*Industr. Engng. Chem.* **40**: 1363.
 HARKINS, W. D., and KIRKWOOD, J. G. (1938).—*J. Chem. Phys.* **6**: 53.
 VAN DER HÖNERT (1932).—*Acad. Sci. Amst.* **35**: 1104.
 MCBAIN, J. W., ROSS, S., BRADY, A. P., ROBINSON, J. V., ABRAMS, I. M., THORNBURN, R. C., and LINDQUIST, C. G. (1944).—National Advisory Committee for Aeronautics, A.R.R. No. 4105.
 NEVILLE, H. A., and HAZELHURST, T. H. (1937).—*J. Phys. Chem.* **41**: 545.
 NUTTING, G. C., HARKINS, W. D., and LONG, F. A. (1940).—*J. Amer. Chem. Soc.* **62**: 1496.
 PATTLE, R. E. (1950).—*J. Soc. Chem. Ind.* **69**: 363.
 RIDEAL, E. K., and SUTHERLAND, K. L. (1952).—*Trans. Faraday Soc.* (in press).
 ROBINSON, J. V., and WOODS, W. W. (1948).—*J. Soc. Chem. Ind.* **67**: 361.
 SCHULMAN, J. H., and TEORELL, T. (1938).—*Trans. Faraday Soc.* **34**: 1337.
 SUTHERLAND, K. L. (1951).—*Rev. Pure and Appl. Chem.* **1**: 35.
 TALMUD, D., and SUCHOWOLOSKAJA, S. (1931).—*Z. Phys. Chem. A* **154**: 277.
 WILSON, R. E., and RIES, E. D. (1923).—*Colloid Symp. Monogr.* **1**: 145-73.

THIOHYDANTOINS

I. THE PREPARATION OF SOME 2-THIOHYDANTOINS FROM AMINO ACIDS AND ACYLAMINO ACIDS

By J. M. SWAN*

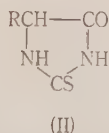
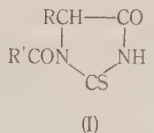
[Manuscript received April 15, 1952]

Summary

A reinvestigation of the stepwise degradation of peptides and proteins by the Schlack and Kumpf (1926) procedure, which involves conversion of the terminal carboxyl residue to a 5-substituted-2-thiohydantoin, has been undertaken. The following amino acids or their acetyl derivatives have been converted by warming with acetic anhydride and ammonium thiocyanate into the corresponding 1-acetyl-2-thiohydantoin: valine, methionine, *isoleucine*, α -aminophenylacetic acid, tryptophane, histidine, and tyrosine. These new 1-acetyl-2-thiohydantoin, as well as others, have been deacetylated by an improved procedure. Certain 1-acyl-5-carbethoxy-2-thiohydantoin on hydrolysis give 2-thiohydantoin only. *N*-acetylthreonine, acetic anhydride, and ammonium thiocyanate produce *5-ethylidene-2-thiohydantoin*, also obtained by the condensation of acetaldehyde with 2-thiohydantoin or its 1-acyl derivatives, or from the interaction of ammonium thiocyanate with 2-phenyl-4-ethylideneoxazol-5-one. This type of rearrangement occurs in other cases.

I. INTRODUCTION

The conversion of an amino acid or an acylamino acid to a 1-acetyl- or 1-acyl-2-thiohydantoin (I) by heating with acetic anhydride and a salt of thiocyanic acid was first observed by Johnson and Nicolet (1911) and by Komatsu (1911). The reaction was subsequently investigated in detail by Johnson and his co-workers.



Schlack and Kumpf (1926) showed that the 1-acyl group in I could be hydrolysed by dilute alkali, the overall reaction being cleavage of the acylamino acid with conversion of the amino acid portion to a 2-thiohydantoin II. If the group R' in I contains further peptide units, then the compound R'CO₂H, from the alkaline fission of I, can in theory be further degraded by this same process, leading to a second 2-thiohydantoin corresponding to the amino acid residue second from the end of the original molecule. Schlack and Kumpf (1926) demonstrated the feasibility of this stepwise degradation with benzoylleucylglycine, benzoylalanylalanine, and benzoyldiglycylglycine, but were unable to

* Biochemistry Unit, Wool Textile Research Laboratory, C.S.I.R.O., Melbourne.

obtain any thiohydantoins from benzoylated globin. The method was used successfully in showing glycine to be the terminal residue of glutathione (Nicolet 1930a). Haurowitz and Vardar (1944) found that, when casein was treated with acetic anhydride and ammonium thiocyanate, the sulphur percentage rose from 0.65 to 1.24 and fell again to 0.92 after 3 hours' treatment with 1N sodium hydroxide. However, Haurowitz, Lisie, and Bursa (1947) and Haurowitz (1950) have claimed that treatment of these substituted proteins with alkali produces 2 moles of thiourea per protein molecule, and no thiohydantoins were isolated.

A further investigation of this technique for the determination of the C-terminal residues, i.e. the carboxyl end-groups, in peptides and proteins has been carried out, and the results are given in this and subsequent papers. In the present work, the list of thiohydantoins derived from amino acids has been further extended. Some observations have also been made on the dilute acid hydrolysis of 1-acyl-5-substituted-2-thiohydantoins, and on a new preparation of 5-alkylidene- and 5-arylidene-2-thiohydantoins from the corresponding 4-alkylidene- or 4-arylidene-2-substituted-oxazol-5-ones. Four of the compounds described herein were reported independently by Kuck *et al.* (1951) after this work had been completed. In addition, Tibbs (1951) has successfully used the thiohydantoin degradation on small quantities of peptides, while Waley and Watson (1951) have applied it to a study of the end-groups of insulin.

II. DISCUSSION

(a) The Preparation of Some Miscellaneous 2-Thiohydantoins

Valine, isoleucine, methionine, α -aminophenylacetic acid, tryptophane, and *N*-acetyl- α -carbethoxyglycine all react smoothly with a solution containing ammonium thiocyanate, acetic anhydride, and 10 per cent. acetic acid to yield respectively 5-isopropyl-, 5-sec.-butyl-, 5-3-methylthiomethyl-, 5-phenyl-, 5-3-indolylmethyl-, and 5-carbethoxy-1-acetyl-2-thiohydantoin. *N*-Phenylacetyl- α -carbethoxyglycine similarly gives 1-phenylacetyl-5-carbethoxy-2-thiohydantoin (I; R = CO₂Et, R' = C₆H₅CH₂). The isoleucine, α -aminophenylacetic acid, and tryptophane derivatives are deacetylated by 2N ammonia or 1N sodium hydroxide to yield the corresponding 5-sec.-butyl-, 5-phenyl-, and 5-3-indolylmethyl-2-thiohydantoin (II; R = C₂H₅CH(CH₃), C₆H₅, and C₈H₆N.CH₂ respectively). In compounds of type II when R is COOH or COOEt, carbon dioxide is apt to be so readily lost that in some instances, described in Section III, decarboxylation has prevented the isolation of the desired thiohydantoin derivative II. Similarly Dains and Krober (1939) isolated rhodanine instead of its 5-carbethoxy-derivative.

Both histidine and *N*-acetylhistidine react on gentle warming with acetic anhydride and ammonium thiocyanate to yield 1-acetyl-5-(1'-acetyl-4-iminazolylmethyl)-2-thiohydantoin, the histidine iminazole ring being simultaneously acetylated. This unstable compound is converted by mild hydrolytic agents (e.g. dilute acetic acid) and by ethanol into a mono-acetyl compound, m.p. 230°C., which ultraviolet absorption measurements indicate to be 5-(1'-acetyl-

4'-iminazolylmethyl)-2-thiohydantoin rather than the isomeric 1-acetyl compound.

The preparation of an *NN'*-diacetylbis-2-thiohydantoin from cystine has been described by Nicolet (1930*b*), but under his conditions, sulphur and 1-acetyl-5-acetylthiomethyl-2-thiohydantoin were also formed. This type of breakdown (cf. Tarbell and Harnish 1951) can be minimized by working at lower temperatures. Refluxing with 1*N* acetic acid converts the non-crystalline thiohydantoin prepared from *NN'*-diacetyllysine to 5- δ -acetamido-*n*-butyl-2-thiohydantoin (II; $R = CH_3CONH(CH_2)_4-$); while the crude thiohydantoin prepared from tyrosine (Johnson and Nicolet 1913; Asahina 1930), on boiling with aqueous ethanol, gave two distinct mono-acetyl derivatives, m.p.'s 144 and 208-209 °C., whose structures remain uncertain.

It is now found that dilute acid, as well as concentrated acid (Johnson and co-workers) and alkali (Schlack and Kumpf 1926), is convenient for the deacylation of 1-acyl-2-thiohydantoins, and may be expected to minimize degradation of a protein or peptide residue attached to the ring. For some 12 different compounds, listed in Table 2, the yields were excellent and the products of high purity, although the time required in some cases was as long as 5 hours. In other cases, refluxing with water was sufficient, and the ready deacylation of glycine thiohydantoins probably explains why Schlack and Kumpf (1926) were able to obtain only 2-thiohydantoin from the direct reaction of diglycylglycine with ammonium thiocyanate and acetic anhydride. It has not been possible to confirm the hydrolysis of 1-acetyl-2-thiohydantoins by cold ethanolic mercuric chloride, reported by Clarke, Johnson, and Robinson (1949, p. 289).

All attempts to prepare thiohydantoins from arginine or serine have been unsuccessful. It is known (Bergmann and Köster 1926) that the action of warm acetic anhydride on mono-acetylarginine gives triacetylanhydroarginine, which on treatment with water gives diacetylurea and DL- β -acetyl-amino- α -piperidone (cf. also Rinderknecht and Niemann 1948) while the serine products were intractable, due probably to polymerization (cf. Dakin and West 1928; Nicolet 1930*c*).

Both acetylaspartic and acetylglutamic acids fail to yield thiohydantoins directly, being converted to cyclic anhydrides which do not react with the ammonium thiocyanate. However, conversion to thiohydantoins is possible by first allowing these anhydrides to react with amines (see Part II of this series, Swan 1952).

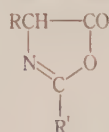
(*b*) Formation of a 2-Thiohydantoin from Threonine

In discussing the mechanism of the thiohydantoin reaction, Johnson and Scott (1913) rejected the earlier theory of Johnson and Nicolet (1911) and suggested that the active intermediate was the oxazol-5-one (III), formed by dehydration of the acylamino acid. Ring cleavage of III by the thiocyanic acid would give the acylisothiocyanate (IV) and thence the 1-acyl-2-thiohydantoin (I).

This mechanism was supported by the demonstration that the oxazolones (III; $R = H$ and CH_3 , $R' = C_6H_5$) react with ammonium thiocyanate in acetic

acid to form the thiohydantoins (I; $R = H$ and CH_3 , $R' = C_6H_5$), and further examples are provided by Clarke, Johnson, and Robinson (1949, pp. 286-7).

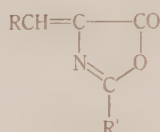
For threonine (β -hydroxy- α -amino-*n*-butyric acid), dehydration of the *N*-acetyl derivative would be expected to give the ethylideneoxazolone (V; $R = R' = CH_3$) (cf. Carter, Handler, and Melville 1911). The further rearrangement to a thiohydantoin in this case was problematical, in view of the claim by



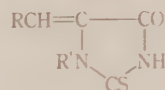
(III)



(IV)



(V)



(VI)

Johnson and Scott (1913) that 4-*unsaturated* oxazolones (V) do not react in this way. However, a small yield of a 2-thiohydantoin was in fact obtained from *N*-acetylthreonine, but proved to be not the expected VI ($R = CH_3$, $R' = CH_3CO$) but the deacetylated compound, 5-ethylidene-2-thiohydantoin (VI; $R = CH_3$, $R' = H$).

This same compound can be prepared independently by the condensation of either 2-thiohydantoin, 1-acetyl-2-thiohydantoin, or 1-benzoyl-2-thiohydantoin with acetaldehyde, and also by heating 2-phenyl-4-ethylideneoxazol-5-one with ammonium thiocyanate in glacial acetic acid. In no case was any of the 1-acyl derivative encountered. Similar deacetylations have been observed during the condensation of benzaldehyde with 1-acetyl-2-thiohydantoin (Asahina 1930; Namjoshi and Dutt 1931; Boyd and Robson 1935) and with 1-benzoyl-2-thiohydantoin (Johnson and Nicolet 1911; Asahina 1930). The conversion of oxazolones of type V to thiohydantoins has been extended to 2-methyl- and 2-phenyl-4-benzylideneoxazol-5-one (V; $R' = CH_3$ and C_6H_5 , $R = C_6H_5$), both of which yield 5-benzylidene-2-thiohydantoin (VI; $R = C_6H_5$, $R' = H$).

III. EXPERIMENTAL

Analyses are by Dr. Weiler and Dr. Strauss, Oxford, and Dr. K. W. Zimmermann, Division of Industrial Chemistry, C.S.I.R.O., Melbourne. Melting points are uncorrected.

(a) Preparation of 1-Acetyl-2-thiohydantoins (I; $R' = CH_3CO$) from Amino Acids

It has been shown (Johnson and Nicolet 1913; Johnson, Hill, and Bailey 1915) that ammonium thiocyanate is much superior to all other salts of thiocyanic acid in the thiohydantoin

TABLE I
PREPARATION OF 1-ACETYL-2-THIOHYDANTOINS

Amino Acid	R in I ($R' = CH_3CO$)	Yield (%)	Melting Point (°C.)	Formula	Found (%)				Calculated (%)			
					C	H	N	S	C	H	N	S
α -Aminophenyl- acetic acid ..	C_6H_5	62	199-201	$C_{11}H_{13}O_4N_2S$	56.7	4.3	11.8	—	56.4	4.3	12.0	—
Isoleucine ..	$CH_3CH(CH_3)CH_2$	65	161-163	$C_{14}H_{19}O_4N_2S$	59.5	6.6	13.7	—	59.4	6.6	13.1	—
Methionine ..	$CH_2SCH_2CH_3$	70	192-193	$C_{12}H_{17}O_4N_2S_2$	41.4	5.2	—	27.5	41.4	5.2	—	27.6
Tryptophan ..	$C_6H_5N.CH_2$	45	172	$C_{14}H_{15}O_4N_3S$	59.0	4.6	14.8	—	58.5	4.6	14.6	—
Valine ..	$(CH_3)_2CH$	—	112	$C_{10}H_{15}O_4N_2S$	47.9	6.0	14.1	—	48.0	6.0	14.0	—

reaction, and it has been used exclusively in the present investigation. The following procedure is typical for the amino acids in Table I, where the results are shown: Valine (1.2 g.)

and ammonium thiocyanate (0.9 g.) were warmed in acetic anhydride containing 10% acetic acid (10 ml.) to give a clear yellow solution, and the solvents were then evaporated under reduced pressure. The solid residue was recrystallized from aqueous ethanol to give colourless needles of the thiohydantoin, m.p. 112 °C.

(b) *1-Acetyl-5-carbethoxy-2-thiohydantoin*

Diethylacetamidomalonate (6.3 g.) (Cerchez 1930; Snyder and Smith 1944) was added to a solution of potassium hydroxide (1.7 g.) in ethanol (25 ml.). A solid soon commenced to separate and after 3 hours the solvent was evaporated, the residue dissolved in water (10 ml.), and the solution acidified with 10N hydrochloric acid (3 ml.). *Ethyl hydrogen acetamidomalonate* (*N*-acetyl- α -carbethoxyglycine) (3 g.) separated on cooling and crystallized from water in short irregular prisms, m.p. 140–141.5 °C. with evolution of carbon dioxide.

Found: C, 45.0; H, 5.8; N, 7.4%.

Calculated for $C_7H_{11}O_5N$: C, 44.4; H, 5.9; N, 7.4%.

A mixture of this acid (2 g.), acetic anhydride (18 ml.), acetic acid (2 ml.), and ammonium thiocyanate (0.9 g.) was warmed to give a clear solution, allowed to stand overnight, and then evaporated to small bulk under reduced pressure. The residue was evaporated twice with ethanol, shaken with water, and the insoluble oil extracted with ether. Evaporation afforded a gum which crystallized on standing. *1-Acetyl-5-carbethoxy-2-thiohydantoin* crystallized from chloroform on addition of light petroleum (40–60 °C.) in white irregular prisms, 0.5 g., m.p. 104–108 °C. After several recrystallizations the m.p. remained constant at 108–109 °C.

Found: C, 41.7; H, 4.1; N, 12.4%.

Calculated for $C_8H_{10}O_4N_2S$: C, 41.8; H, 4.4; N, 12.2%.

In an attempt to prepare the 5-carbamido analogue, a solution of the above ethyl hydrogen acetamidomalonate (5 g.) in concentrated ammonia (20 ml.) was allowed to stand at room temperature for 3 days, the solution evaporated in a vacuum, the residue acidified with 6N hydrochloric acid (4.4 ml.), and again evaporated. Extraction with acetone and evaporation of the solvent gave a clear oil which slowly solidified (2.8 g.). Crystallization from a small amount of hot ethanol yielded 0.8 g. of acetylglutamine amide, m.p. 137–137.5 °C. after 4 recrystallizations from ethanol, undepressed by admixture with an authentic specimen.

Found: C, 41.6; H, 7.2%.

Calculated for $C_4H_8O_2N_2$: C, 41.4; H, 6.9%.

Evaporation of the first ethanolic mother liquor gave a gum (2 g.) which after the usual thiohydantoin reaction yielded 0.5 g. of 1-acetyl-2-thiohydantoin, pale yellow, thick, hexagonal plates from ethyl acetate, m.p. 179 °C.

Found: C, 38.1; H, 3.8; S, 20.3%.

Calculated for $C_5H_8O_2N_2S$: C, 38.0; H, 3.8; S, 20.3%.

(c) *1-Phenylacetyl-5-carbethoxy-2-thiohydantoin*

Ethyl hydrogen phenylacetaminomalonate (1.0 g.) (Clarke, Johnson, and Robinson 1949, p. 777), ammonium thiocyanate (0.5 g.), acetic anhydride (9 ml.), and acetic acid (1 ml.) were heated on the steam-bath for 30 minutes. The product, on working up in the usual manner, weighed 1.1 g., m.p. 88–100 °C. On crystallization from aqueous ethanol the m.p. rose to 110 °C.

Found: C, 55.1; H, 4.2; N, 9.0%.

Calculated for $C_{14}H_{14}O_4N_2S$: C, 54.9; H, 4.6; N, 9.2%.

(d) *Phenylacetamidomalononic Acid*

The diethyl ester (2 g.) (Clarke, Johnson, and Robinson 1949, p. 880) was allowed to react for 12 hours at room temperature with a solution of potassium hydroxide (0.76 g.) in ethanol (4 ml.). A solid separated, was filtered, washed with a little ethanol, dissolved in water, and the solution treated with charcoal and then acidified with 6N sulphuric acid (1 ml.). A large crop of

colourless needles separated slowly (0.5 g.). The substance lost carbon dioxide on heating and melted at 142 °C., the m.p. of phenaceturic acid. When heated rapidly a mixture of the two melted at 127 °C.

Found: N, 5.7%.

Calculated for $C_{11}H_{11}O_5N$: N, 5.9%.

The original ethanolic filtrate on acidification and extraction with ether yielded the half-ester (Clarke, Johnson, and Robinson 1949, p. 777), m.p. 106–107 °C., undepressed by mixing with an authentic sample.

(e) *1-Acetyl-5-(1'-acetyl-4'-iminazolylmethyl)-2-thiohydantoin*

(i) Mono-acetylhistidine hydrate (2.1 g.) (Bergmann and Zervas 1928; Neuberger 1938) was shaken for 12 hours with a mixture of acetic anhydride (18 ml.), acetic acid (2 ml.), and ammonium thiocyanate (0.84 g.), and the insoluble *product* filtered, 1.0 g., m.p. 186–188 °C., and recrystallized from ethanol, m.p. 187–188 °C.

Found: C, 47.2; H, 4.3; S, 11.3%.

Calculated for $C_{11}H_{12}O_5N_4S$: C, 47.1; H, 4.3; S, 11.4%.

Ultraviolet absorption in ethanol: λ_{max} 275 m μ , ϵ = 13,900; λ_{min} 248 m μ , ϵ = 7200.

(ii) DL-Histidine monohydrochloride dihydrate (3 g.), ammonium thiocyanate (1.0 g.), anhydrous sodium acetate (1.0 g.), acetic anhydride (9 ml.), and acetic acid (1 ml.) were heated together at 50 °C. for 5 minutes until a suspension of sodium chloride only was present in the liquid. After standing for a further 30 minutes the solvents were evaporated under reduced pressure; the residue was evaporated twice more with water and then filtered, 1.8 g. (48.5%), m.p. 183–184 °C.

(f) *5-(1'-Acetyl-4'-iminazolylmethyl)-2-thiohydantoin*

(i) The diacetyl compound above was allowed to stand in cold ethanol for 48 hours. The solid slowly dissolved and evaporation gave a pinkish fluffy solid. This was purified by redissolving in ethanol and precipitating with chloroform. The first dark red oily fractions were discarded; further dilution with cooling gave a very small yield of globular lumps of a cream coloured solid, m.p. 228–229 °C. (decomp.). Attempts to purify this material further led to more decomposition.

(ii) The diacetyl compound (0.4 g.) was allowed to stand for 3 hours with 1N sodium hydroxide (2 ml.). Acidification with 6N sulphuric acid gave 0.2 g. of a crystalline deposit, m.p. 226 °C., which recrystallized from 50% alcohol in yellow needles, m.p. 230 °C. (decomp.), unaltered by admixture with the material above having m.p. 228–229 °C. (decomp.).

Found: S, 13.4%.

Calculated for $C_9H_{10}O_2N_4S$: S, 13.5%.

(iii) The diacetyl compound (0.5 g.) was refluxed with water or 1N acetic acid for 3 hours. Evaporation gave a gum from which 0.04 g. of a cream coloured solid, m.p. 227 °C. (decomp.), was eventually separated by precipitation from ethanol solution with chloroform. Ultraviolet absorption in ethanol: λ_{max} 268 m μ , ϵ = 11,700; λ_{min} 248 m μ , ϵ = 2500. A hypsochromic shift of the absorption maximum by about 10 m μ , together with a smaller extinction coefficient, corresponds to removal of the thiohydantoin ring-acetyl group. Compare data on simpler thiohydantoins given by Asahina (1930) and by Clarke, Johnson, and Robinson (1949, p. 287).

(g) *$\alpha\alpha'$ -Bis-(1-acetyl-2-thiohydantoyl-5-yl)dimethyl Disulphide*

NN'-Diacetylcystine (16 g.), ammonium thiocyanate (7.5 g.), acetic anhydride (90 ml.), and acetic acid (10 ml.) were shaken together at room temperature. A white solid soon commenced to separate, was filtered after 24 hours, washed with acetic acid, and then with water. The product weighed 4 g. and melted with frothing to a yellow liquid at 207 °C. (Nicolet 1930b gives m.p. 208 °C.) In other preparations melting points as high as 220 °C. were recorded. On

crystallization from hot dioxane by the addition of water the melting point fell to 198–202 °C. but the product gave a satisfactory analysis.

Found: C, 35.5; H, 3.4; N, 14.5; S, 31.5%.

Calculated for $C_{12}H_{14}O_4N_4S_2$: C, 36.0; H, 3.7; N, 13.9; S, 31.5%.

(h) *5-δ-Acetamido-n-butyl-2-thiohydantoin*

NN'-Diacetyllysine (2.2 g.) (Gordon, Martin, and Syngé 1943) was shaken at room temperature for 18 hours with acetic anhydride (18 ml.), acetic acid (2 ml.), and ammonium thiocyanate (1.0 g.). The clear solution was evaporated under reduced pressure to half-volume, diluted with water (100 ml.), and refluxed for 3 hours. Evaporation gave a yellow oil which crystallized on shaking with water. Yield 0.8 g., m.p. 194 °C. after sintering at 189 °C., unchanged on further crystallization from water.

Found: N, 18.6%.

Calculated for $C_9H_{15}O_2N_3S$: N, 18.3%.

Attempts to obtain the intermediate 1-acetyl derivative in crystalline form were unsuccessful. Hydrolysis of the oil with 1*N* sodium hydroxide gave only a very low yield of the thiohydantoin, m.p. 191–192 °C. after sintering at 187 °C. (50 mg. from 0.9 g. of diacetyllysine).

(i) *Acetylthiohydantoins from Tyrosine*

Tyrosine (0.9 g.), ammonium thiocyanate (0.4 g.), acetic anhydride (9 ml.), and acetic acid (1 ml.) were heated together on a steam-bath for 1 hour and the solvents were then evaporated. The residual oil was evaporated twice more with water and then crystallized five times from 25% aqueous ethanol. The product separated in colourless needles, m.p. 144 °C.

Found: C, 54.7; H, 4.9; N, 10.5%.

Calculated for $C_{12}H_{12}O_3N_2S$: C, 54.5; H, 4.6; N, 10.6%.

Ultraviolet absorption in ethanol: λ_{max} 279 m μ , ϵ =18,300; λ_{min} 249 m μ , ϵ =6000. When this experiment was repeated, the above compound was not found, but an isomeric substance was isolated which crystallized from aqueous ethanol in glistening needles, m.p. 208–209 °C.

Found: C, 54.5; H, 4.7; S, 13.4%.

Calculated for $C_{12}H_{12}O_3N_2S$: C, 54.5; H, 4.6; S, 12.1%.

Ultraviolet absorption in ethanol: λ_{max} 279 m μ , ϵ =15,800; λ_{min} 249 m μ , ϵ =3750. These two compounds are probably the *O*-acetyl and 1-*N*-acetyl derivatives of 5-*p*-hydroxybenzyl-2-thiohydantoin but unfortunately no distinction could be made between them. The ultraviolet absorption curves were very similar, and both compounds failed to give a colour with ferric chloride or more than a transitory red colour when treated in alkali with diazotized-*p*-nitro-aniline.

(j) *Hydrolysis of 1-Acyl-2-thiohydantoins*

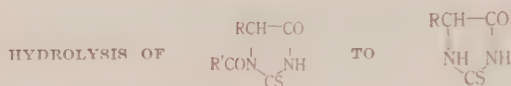
The results obtained from a wide range of compounds are shown in Table 2. The products were identified either by analysis or by mixed melting point, and were crystallized either from water or aqueous ethanol. References to earlier methods for the preparation of the compounds involved are given in the footnote to Table 2.

(k) *Attempted Hydrolysis using Cold Mercuric Chloride*

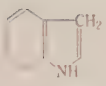
It has been claimed (Clarke, Johnson, and Robinson 1949, p. 289) that the action of cold ethanolic mercuric chloride on L-1-acetyl-5-isobutyl-2-thiohydantoin, followed by hydrogen sulphide to remove mercury, gives the deacetylated racemic-5-isobutyl-2-thiohydantoin. Repetition of this work with close adherence to the method as described has given only unchanged starting material. There is some uncertainty as to the amount of mercuric chloride used by the Squibb workers, but the use of both 5 moles and 1 mole gave the same result. The reaction of the thiohydantoin, and later, the addition of hydrogen sulphide, liberates free hydrochloric acid from the mercuric chloride, and undue heating of the acid solutions probably effected the

deacetylation. Similar experiments with another thiohydantoin also failed. Thus 1-acetyl-5-methyl-2-thiohydantoin (0.5 g.) in methanol (5 ml.) was added to a solution of mercuric chloride (1 g.) in 50% aqueous methanol (10 ml.). A thick white precipitate appeared and after standing

TABLE 2



Methods: *A*, reflux with water; *B*, reflux with 1*N* acetic acid; *C*, heat to 70 °C. with 2*N* ammonia, cool, and acidify

R	R'	Method	Reaction Time (hr.)	Melting Point of Product (°C.)	Yield (%)
H	CH ₃ ^(a)	<i>A</i>	1.5	227 (decomp.) ^(a)	82
		<i>B</i>	1	"	95
H	C ₆ H ₅ ^(a)	<i>A</i>	1	"	75
CH ₃	CH ₃ ^(a)	<i>B</i> ^(b)	2	159–160 ^(a)	98
(CH ₃) ₂ CH	CH ₃ ^(c, d)	<i>B</i>	2	139–140 ^(e, f)	96
		<i>C</i>	—	"	50
(CH ₃) ₂ CHCH ₂	CH ₃ ^(d, g)	<i>B</i>	5 ^(h)	172–173 ^(g)	40
CH ₃ SCH ₂ CH ₂	CH ₃ ^(c, d)	<i>A</i>	4	148–149 ^(e, f, j)	40
		<i>C</i>	—	"	92
C ₆ H ₅	CH ₃ ^(c)	<i>B</i> ^(k)	4	223–225 ^(c, l)	—
		<i>C</i>	—	"	70
	CH ₃ ^(c)	<i>B</i>	5 ^(h)	190–191 ^(e, m)	95
		<i>C</i>	—	"	50
H ₂ NCOCH ₃	CH ₃ ^(a, d)	<i>B</i>	6 ^(h)	246 ^(a)	—
H ₂ NCOCH ₂ CH ₃	CH ₃ ⁽ⁿ⁾	<i>B</i>	4	190–191 ^(a, n, o)	—
		<i>C</i>	—	"	45
EtO ₂ C	CH ₃ ^(c)	<i>B</i>	2	227 (decomp.) ^(p)	—
EtO ₂ C	C ₆ H ₅ (^c)CH ₃ ^(c)	<i>B</i>	2	"	—
		<i>C</i>	—	"	—

^(a) Johnson (1913). ^(b) Method *A* was unsuccessful. ^(c) Described in this paper. ^(d) Kuck *et al.* (1951). ^(e) Jackman *et al.* (1951). ^(f) Found: C, 45.8; H, 6.1%. Calc. for C₈H₁₀ON₂S: C, 45.5; H, 6.4%. ^(g) Schlack and Kumpf (1926). ^(h) Shorter reaction time was insufficient. ⁽ⁱ⁾ Burger and Cloyne (1928). ^(j) Found: C, 38.2; H, 5.2%. Calc. for C₈H₁₀ON₂S₂: C, 37.9; H, 5.3%. ^(k) Dioxane was added. ^(l) Found: C, 56.3; H, 4.4; N, 14.1%. Calc. for C₈H₈ON₂S: C, 56.3; H, 4.2; N, 14.6%. ^(m) Found: C, 59.2; H, 4.6; N, 16.8%. Calc. for C₁₂H₁₁ON₂S: C, 58.8; H, 4.5; N, 17.1%. ⁽ⁿ⁾ Nicolet (1930*d*). ^(o) Found: C, 38.8; H, 4.8; N, 22.5%. Calc. for C₆H₉O₃N₂S: C, 38.5; H, 4.9; N, 22.5%. ^(p) Product isolated was 2-thiohydantoin. Found: N, 23.5%. Calc. for C₃H₄ON₂S: N, 24.1%.

for 1.5 hours hydrogen sulphide was passed into the mixture. Filtration and evaporation under reduced pressure gave unchanged starting material, m.p. 168 °C. after one recrystallization. The same result was obtained using ethanol solutions and a fivefold excess of mercuric chloride, and allowing the mixture to stand for 24 hours.

(l) *5-Ethylidene-2-thiohydantoin*

(i) *N*-Acetylthreonine (1.2 g.) (prepared from threonine as described for *N*-acetylserine by Syngé 1939) was treated in the cold with acetic anhydride (9 ml.), acetic acid (1 ml.), and ammonium thiocyanate (0.75 g.) for 18 hours; the mixture was then diluted with water and evaporated in a vacuum at 50 °C. After a second evaporation with water the residue was extracted into ether, evaporation of which gave a yellow oil which partly crystallized on keeping. This was boiled with chloroform (50 ml.) and the insoluble crystalline material filtered, 0.1 g., m.p. 260–265 °C. (decomp.). The filtrate failed to yield further crystalline products. The solid was crystallized twice from acetic acid and identified as *5-ethylidene-2-thiohydantoin*. The substance crystallized in flat needles and decomposed slowly between 250 and 259 °C. to a black tar with preliminary darkening at 245 °C. (slow heating) or at 263–264 °C. (rapid heating).

Found: C, 42.2; H, 4.9; S, 22.5%.

Calculated for $C_8H_9ON_2S$: C, 42.2; H, 4.3; S, 22.5%.

Initial attempts to prepare this compound by the condensation of acetaldehyde with 1-acetyl-2-thiohydantoin were unsuccessful. Experiments using acetic anhydride containing sodium acetate, lead acetate, or sulphuric acid; or pyridine containing piperidine (Boyd and Robson 1935) as condensing medium, with wide variation of reaction temperature, gave rise to tars and intractable products. The compound was eventually obtained in four different ways as described below.

(ii) Acetic acid (5 ml.), anhydrous lead acetate (0.8 g.), 1-acetyl-2-thiohydantoin (0.8 g.), and acetaldehyde (3 ml.) were heated in a sealed tube at 90 °C. for 3 hours. Crystals soon appeared in the hot solution and cooling and filtering gave 0.4 g. of *5-ethylidene-2-thiohydantoin*, flat colourless needles from acetic acid, m.p. 263–265 °C. (decomp.), with preliminary darkening at 245 °C., alone or when admixed with the product from *N*-acetylthreonine above.

Found: C, 42.5; H, 4.3; S, 21.3%.

A further 0.1 g. was obtained from the original acetic acid mother liquor. Total yield 70%. The same product was obtained in 52% yield when the mixture was heated slowly for 1.5 hours from 60–90 °C. and then held at 80 °C. for a further 1.5 hours. When 1-benzoyl-2-thiohydantoin, or 2-thiohydantoin itself was used in place of 1-acetyl-2-thiohydantoin, the yields were 56 and 68% respectively.

(iii) 2-Phenyl-4-ethylideneoxazol-5-one (0.5 g.) (Finar and Libman 1949), acetic anhydride (5 ml.), and ammonium thiocyanate (0.5 g.) were heated in a sealed tube for 6 hours at 105 °C. Cooling and filtering gave 135 mg. of *5-ethylidene-2-thiohydantoin*, which after crystallization from acetic acid had m.p. 258–260 °C. (decomp.).

Found: C, 42.3; H, 4.5; N, 19.9; S, 23.0%.

(m) *5-Benzylidene-2-thiohydantoin*

(i) 2-Methyl-4-benzylideneoxazol-5-one (0.5 g.) (Herbst and Shemin 1943), ammonium thiocyanate (0.5 g.), and acetic acid (5 ml.) were heated together in a sealed tube at 105 °C. until crystals commenced to separate after 4.5 hours from the hot solution. The mixture was then cooled, filtered, and the product, 0.17 g., m.p. 257 °C. (decomp.), recrystallized from acetic acid as yellow needles, m.p. 258 °C. (decomp.). Johnson and Nicolet (1911) report m.p. 258 °C. (decomp.) and Asahina (1930) reports 260 °C. (decomp.).

Found: C, 58.9; H, 3.9%.

Calculated for $C_{10}H_8ON_2S$: C, 58.8; H, 3.9%.

(ii) 2-Phenyl-4-benzylideneoxazol-5-one (0.5 g.), acetic acid (5 ml.), and ammonium thiocyanate (0.5 g.) were heated together in a sealed tube at 105 °C. for 24 hours. Cooling gave 0.1 g. of dark brown crystals, m.p. 255–257 °C. after recrystallization from acetic acid, undepressed by admixture with authentic *5-benzylidene-2-thiohydantoin*.

IV. ACKNOWLEDGMENT

The author is grateful to Dr. J. A. Friend of this Laboratory for the measurement of the absorption spectra.

V. REFERENCES

- ASAHINA, T. (1930).—*Bull. Chem. Soc. Japan* **5**: 354.
BARGER, G., and CLOYNE, F. P. (1928).—*Biochem. J.* **22**: 1417.
BERGMANN, M., and KÖSTER, H. (1926).—*Hoppe-Seyl. Z.* **159**: 179.
BERGMANN, M., and ZERVAS, L. (1928).—*Biochem. Z.* **203**: 284.
BOYD, W. J., and ROBSON, W. (1935).—*Biochem. J.* **29**: 542.
CARTER, H. E., HANDLER, P., and MELVILLE, D. B. (1911).—*J. Biol. Chem.* **129**: 364.
CERCHEZ, V. (1930).—*Bull. Soc. Chim.* **47**: 1282.
CLARKE, H. T., JOHNSON, J. R., and ROBINSON, R. (Editors) (1949).—"The Chemistry of Penicillin." (Princeton Univ. Press.)
DAINS, F. B., and KROBER, O. A. (1939).—*J. Amer. Chem. Soc.* **61**: 1830.
DAKIN, H. D., and WEST, R. (1928).—*J. Biol. Chem.* **78**: 745.
FINAR, I. L., and LIBMAN, D. D. (1949).—*J. Chem. Soc.* **1949**: 2726.
GORDON, A. H., MARTIN, A. J. P., and SYNGE, R. L. M. (1943).—*Biochem. J.* **37**: 79.
HAUROWITZ, F. (1950).—"Chemistry and Biology of Proteins." p. 109. (Academic Press: New York.)
HAUROWITZ, F., LISIE, S., and BURSA, F. (1947).—Abstr. 112th Meeting Amer. Chem. Soc. 29C.
HAUROWITZ, F., and VARDAR, M. (1944).—*Arch. Soc. Turq. Sci. Phys. Nat.* **11**: 33.
HERBST, R. M., and SHEMIN, D. (1943).—*Organic Synth.* (Coll. Vol.) **2**: 1.
JACKMAN, M., KLENK, M. M., FISHBURN, B., TULLER, B. F., and ARCHER, S. (1949).—*J. Amer. Chem. Soc.* **70**: 2884.
JOHNSON, T. B. (1913).—*Amer. Chem. J.* **49**: 68.
JOHNSON, T. B., HILL, A. J., and BAILEY, B. H. (1915).—*J. Amer. Chem. Soc.* **37**: 2415.
JOHNSON, T. B., and NICOLET, B. H. (1911).—*J. Amer. Chem. Soc.* **33**: 1973.
JOHNSON, T. B., and NICOLET, B. H. (1913).—*Amer. Chem. J.* **49**: 197.
JOHNSON, T. B., and SCOTT, W. M. (1913).—*J. Amer. Chem. Soc.* **35**: 1136.
KUCK, J. F. R., HERDA, J. J., KOVAC, W. E., and KARABINOS, J. V. (1951).—*J. Amer. Chem. Soc.* **73**: 5470.
KOMATSU, S. (1911).—*Mem. Coll. Sci. Engng. Kyoto* **3**: 1. (*Chem. Zbl.* **11**: 537.)
NAMJOSHI, V. G., and DUTT, S. (1931).—*J. Indian Chem. Soc.* **8**: 241.
NEUBERGER, A. (1938).—*J. Biol. Chem.* **32**: 1452.
NICOLET, B. H. (1930a).—*Science* **71**: 589; *J. Biol. Chem.* **88**: 389.
NICOLET, B. H. (1930b).—*J. Biol. Chem.* **88**: 395.
NICOLET, B. H. (1930c).—*J. Biol. Chem.* **88**: 403.
NICOLET, B. H. (1930d).—*J. Amer. Chem. Soc.* **52**: 1192.
RINDERKNECHT, H., and NIEMANN, C. (1948).—*J. Amer. Chem. Soc.* **70**: 2605.
SCHLACK, P., and KUMPE, W. (1926).—*Hoppe-Seyl. Z.* **154**: 125.
SNYDER, H. R., and SMITH, C. W. (1944).—*J. Amer. Chem. Soc.* **66**: 350.
SWAN, J. M. (1952).—*Aust. J. Sci. Res. A* **5**: 721.
SYNGE, R. L. M. (1939).—*Biochem. J.* **33**: 1924.
TARBELL, D. S., and HARNISH, D. P. (1951).—*Chem. Rev.* **49**: 11.
TIBBS, J. (1951).—*Nature* **168**: 910.
WALEY, S. G., and WATSON, J. (1951).—*J. Chem. Soc.* **1951**: 2394.

THIOHYDANTOINS

II. THIOHYDANTOINS DERIVED FROM ASPARTIC AND GLUTAMIC ACIDS

By J. M. SWAN*

[Manuscript received April 15, 1952]

Summary

N-Acyl derivatives of the dicarboxylic acids, aspartic and glutamic acids, yield the corresponding anhydrides when heated with acetic anhydride in the presence or absence of ammonium thiocyanate. In general these anhydrides form isomeric pairs of amides and anilides, and only one of each pair can be converted into the corresponding 2-thiohydantoin. The anhydrides from *N*-phenylacetyl- and *N*-*p*-nitrobenzoylglutamic acids, and *N*-acetylaspargic acid have been regarded hitherto as oxazolones.

I. INTRODUCTION

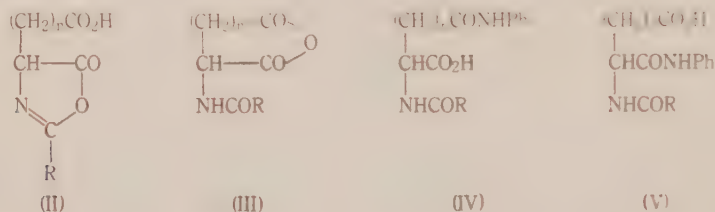
The preceding paper (Part I of this series, Swan 1952*a*) dealt with the preparation of some 1-acyl-5-substituted-2-thiohydantoin. This reaction between an amino acid and acetic anhydride and ammonium thiocyanate is known to fail when applied to aspartic and glutamic acids (Johnson and Guest 1912*a*), and this imposes a limitation on the thiohydantoin method for the determination of the C-terminal residues (carboxyl end-groups) in peptides and proteins. Several acyl derivatives of glutamic acid, and one of aspartic acid, have been studied more fully in the hope of finding a satisfactory means for applying the degradation to peptide chains terminating in these residues. A preliminary account of this work, dealing more particularly with the anhydrides of *N*-acetylaspargic and *N*-acetylglutamic acids, has been given elsewhere (Swan 1952*b*).

II. RESULTS

The acids $\text{HO}_2\text{C}(\text{CH}_2)_n\text{CH}(\text{NHCOR})\text{CO}_2\text{H}$ (I) on dehydration can theoretically yield the oxazolone (II) or the anhydride (III), which with aniline could give the anilides (IV and V). It is now found that acetyl-DL-aspartic acid (I; $\text{R}=\text{CH}_3$, $n=1$) interacts with boiling acetic anhydride, in the presence or absence of ammonium thiocyanate, to give III ($\text{R}=\text{CH}_3$, $n=1$). This explains the failure of the 2-thiohydantoin reaction when applied to aspartic acid (Johnson and Guest 1912*a*). The anhydride (III; $\text{R}=\text{CH}_3$, $n=1$) has already been made and recognized as such by Bergmann, Stern, and Witte (1926), who heated aspartic acid itself with acetic anhydride, but the same product described by Harington and Overhoff (1933) without reference to the earlier paper was thought by them to be II ($\text{R}=\text{CH}_3$, $n=1$).

* Biochemistry Unit, Wool Textile Research Laboratory, C.S.I.R.O., Melbourne.

The anhydride structure is now proved by converting the compound into a mixture of the anilides (IV and V) of which only the β -anilide (IV; $R=CH_3$, $n=1$) yields a 2-thiohydantoin. The product from the reaction of the anhydride with ammonia likewise gives a 2-thiohydantoin, identical with that prepared directly from asparagine (Johnson and Guest 1912*b*; Johnson 1913; Johnson and Nicolet 1913). This structure for acetylaspargic anhydride has also been confirmed recently by Barker (1951)



Thierfelder (1921) recognized the diagnostic value of the thiohydantoin reaction for deciding alternative structures in the glutamic acid series, and Nicolet (1930) showed by experiments similar to the above that the dehydration of acetylglutamic acid likewise gives an anhydride (III; $R=CH_3$, $n=2$). This work is confirmed here.

In two recent papers, the oxazolone structure has been ascribed to an acylglutamic anhydride. Baker and Jones (1951) synthesized an optically inactive compound thought to be II ($R=C_6H_5CH_2$, $n=2$) from *N*-phenylacetyl-L-glutamic acid, and King and Spensley (1950) prepared a racemic anhydride from *p*-nitrobenzoyl-L-glutamic acid and formulated it as II ($R=p-O_2N.C_6H_4$, $n=2$). Both these compounds are shown here to be true anhydrides (III; $n=2$, $R=C_6H_5CH_2$, and $p-O_2N.C_6H_4$ respectively). The phenylacetyl compound with aniline gives two anilides, of which only one (the γ anilide) can be converted to a 2-thiohydantoin. In a preliminary communication (Swan 1952*b*) it was claimed that 'phenylacetyl-DL-glutamic acid α -amide', m.p. 144-146°C., was actually the γ -isomer. However, the results of J. A. King, F. H. McMillan, and J. D. Genzer, which were communicated to the author in advance of publication through the courtesy of Dr. J. A. King, indicate that the amide is a mixture in which the α -isomer, m.p. 148-149°C., predominates. Repetition of the original thiodantoin reaction has given support to this finding, the yield of 1-phenylacetyl-2-thiohydantoin-5- β -propionamide, which is formed from the γ -amide present in the mixture, being never greater than 10 per cent. The structure given by Baker and Jones to 'phenylacetyl-DL-glutamic acid α -hydrazide' has been shown to be correct with its conversion to the α -anilide, while the γ -hydrazide has also been isolated from the reaction mixture.

Similar results were obtained in an investigation of "2-*p*-nitrobenzoyl-4-2'-carboxyethyloxazol-5-one". This *dehydro-N-p*-nitrobenzoylglutamic acid gives two amides and two anilides, of which only one in each pair can be converted to a 2-thiohydantoin. The structures of King and Spensley's *p*-nitrobenzoyl-DL-glutamic acid α -amide and α -benzyl-*N-p*-nitrobenzoyl-DL-glutamate have been confirmed by their non-reactivity in the thiohydantoin reaction.

The above anhydrides all failed to give thiohydantoin, as did also acetyl-, phenylacetyl-, *p*-nitrobenzoyl-, carbobenzyloxy-, or acetylphenylalanyl-L-glutamic acid, when treated with the usual reagents. Since reaction of the anhydrides of acylaspartic or acylglutamic acids with aniline gives in general a mixture of isomers, it would appear that conversion to the anilide only partly solves the problem of applying the thiohydantoin degradation to these acids.

III. EXPERIMENTAL

Analyses are by Dr. K. W. Zimmermann, Division of Industrial Chemistry, C.S.I.R.O. Melting points are uncorrected.

(a) *Acetyl-DL-aspartic Acid*

DL-Aspartic acid (13.3 g., 0.1 mole) was dissolved in 6N sodium hydroxide (34 ml., 0.2 mole) and the solution shaken with acetic anhydride (15 ml.) with cooling in ice. When all the anhydride had dissolved, 10N hydrochloric acid (19 ml.) was added, and the mixture was extracted continuously with ethyl acetate. Evaporation of the solvent gave an oil which soon solidified, and the product was crystallized from ethyl acetate, 5 g., m.p. 140–141 °C. The substance was sparingly soluble in chloroform, ethyl acetate, benzene, and ether, soluble in hot acetone and dioxan, and readily soluble in water and ethanol. Harington and Overhoff (1933) report m.p. 142–143 °C. for a product derived from L-aspartic acid, and having $[\alpha]_{5461}^{20} +5.9^\circ$ in water.

Found: C, 41.2; H, 5.2; N, 8.7%.

Calculated for $C_6H_9O_5N$: C, 41.2; H, 5.2; N, 8.0%.

(b) *Acetyl-DL-aspartic Anhydride*

Acetyl-DL-aspartic acid (0.5 g.) was warmed with acetic anhydride (4 ml.) to give a clear solution. Evaporation to small bulk and cooling gave copious needles of the product (0.3 g.) which was recrystallized from acetic anhydride, and showed m.p. 143–144 °C., depressed to 125–130 °C. on admixture with acetyl-DL-aspartic acid, m.p. 140–141 °C. Bergmann, Stern, and Witte (1926), who prepared this substance by boiling L-aspartic acid with acetic anhydride, report m.p. 141 °C.; Harington and Overhoff (1933) give m.p. 145–146 °C.

Found: N, 8.8%.

Calculated for $C_6H_7O_4N$: N, 8.9%.

The anhydride dissolved only slowly in 2N ammonium bicarbonate, and reacted exothermically with aniline.

(c) *1-Acetyl-2-thiohydantoyl-5- α -acetanilide*

Acetyl-DL-aspartic anhydride (from 1.0 g. of the acid) was treated with aniline (1 ml.). The resultant solid, on application of the thiohydantoin reaction (see preceding paper) yielded 0.37 g. of product, m.p. 242 °C., unchanged on crystallization from ethyl acetate containing a little benzene.

Found: C, 53.4; H, 4.5; N, 14.3%.

Calculated for $C_{13}H_{13}O_3N_3S$: C, 53.6; H, 4.5; N, 14.4%.

(d) α - and β -Anilides of *Acetyl-DL-aspartic Acid*

Acetyl-DL-aspartic anhydride (5 g.) was dissolved in hot ethyl acetate (20 ml.) and treated with aniline (3.2 ml., 1 mole). After 2 hours the crystalline precipitate was filtered (5 g.) and digested with boiling water (50 ml.). Cooling and filtering gave 3.7 g. of product, m.p. 184–191 °C., raised to 202 °C. on recrystallization from methanol by the addition of ether. This substance was recovered unchanged from an attempted thiohydantoin reaction, and gave analytical figures for *acetyl-DL-aspartic acid α -anilide*.

Found: C, 57.8; H, 5.9; N, 10.9%.

Calculated for $C_{12}H_{14}O_4N_2$: C, 57.6; H, 5.6; N, 11.2%.

Evaporation of the aqueous filtrate above gave 1.2 g. of solid, m.p. 163–167 °C. which crystallized in needles from a small volume of acetone by the addition of ether, m.p. 178 °C. (0.5 g.) unaltered by further crystallizations. When mixed with the α -anilide above, the m.p. was 172 °C. Analysis indicated that this was the β -anilide.

Found: N, 11.0, 11.3%.

The usual thiohydantoin reaction gave 1-acetyl-2-thiohydantoyl-5- α -acetanilide, m.p. 242 °C., unchanged on admixture with the substance described above.

(e) *1-Acetyl-2-thiohydantoyl-5- α -acetamide*

Acetyl-DL-aspartic anhydride (0.9 g.) was treated with an excess of dry ammonia in dioxane. After evaporation to dryness the residual oil was converted without further purification to a thiohydantoin. Yield 0.17 g., m.p. 216–217 °C., raised to 219 °C. on admixture with authentic material, m.p. 223–224 °C., prepared from asparagine (Johnson and Guest 1912; Johnson, 1913; Johnson and Nicolet 1913).

(f) *Acetyl-DL-glutamic Acid γ -Anilide*

Acetyl-L-glutamic acid (2 g.) (Bergmann and Zervas 1928; Synge 1939) was heated with acetic anhydride (10 ml.) on the steam-bath to give a clear solution, and the mixture was then evaporated under reduced pressure. The residue was heated with aniline (1 ml.) and allowed to stand for several hours. The product crystallized on boiling with benzene and was filtered (2.3 g.) and recrystallized several times from a mixture of acetone and ether. This was the γ -anilide, m.p. 161–163 °C.

Found: C, 59.5, 59.4; H, 6.2, 6.6; N, 10.5, 10.6%.

Calculated for $C_{13}H_{16}O_4N_2$: C, 59.1; H, 6.1; N, 10.6%.

Application of the thiohydantoin reaction gave 1-acetyl-2-thiohydantoyl-5- γ -propanamide, m.p. 197 °C., showing no m.p. depression with an authentic specimen (Nicolet 1930).

(g) *Phenylacetyl-DL-glutamic Anhydride*

This substance was prepared by Baker and Jones (1951), but was tentatively described by them as "2-benzyl-4-2'-carboxyethyloxazol-5-one". Their method was used here without modification. The substance was only slowly soluble in 2N ammonium bicarbonate and failed to give rise to a thiohydantoin when heated with ammonium thiocyanate in either acetic anhydride or acetic acid.

(h) *α - and γ -Anilides of Phenylacetyl-DL-glutamic Acid*

The above anhydride (3.5 g.) was dissolved in ethyl acetate (20 ml.) and aniline (1.5 ml.) added. A thick solid soon appeared and was filtered after 2 hours, 3.1 g., m.p. 160–165 °C. The mother liquor was evaporated, the residual oil shaken with 2N hydrochloric acid, and the solid filtered, 0.4 g., m.p. 120–150 °C., raised to 159 °C. on two recrystallizations from ethanol. This proved to be phenylacetyl-DL-glutamic acid γ -anilide (for conversion to a thiohydantoin, see below).

Found: C, 67.2; H, 5.8; N, 8.2%.

Calculated for $C_{19}H_{20}O_4N_2$: C, 67.0; H, 5.9; N, 8.2%.

Attempts to purify the main product by crystallization from water, 25% ethanol or ethyl acetate were not promising, but the pure α -anilide was readily obtained by a fractional extraction procedure (cf. Le Quesne and Young 1950). The solid (1.5 g.) was dissolved in ethyl acetate (120 ml.) and extracted five times with 10 ml. portions of a solution of sodium carbonate (0.24 g.) in water (50 ml.). Acidification gave 0.2 g. of product from each extract, the melting points being as follows: (i) 162–164 °C., (ii) 164–165 °C., (iii) 177–180 °C., (iv) 188–190 °C., (v) 192–193 °C. A sixth fraction (0.35 g., m.p. 192–193 °C.) was obtained by evaporating the ethyl acetate solution to 10 ml. and cooling. Fractions (iv), (v), and (vi) were substantially

pure *phenylacetyl-DL-glutamic acid α -anilide*, m.p. 195 °C. after recrystallization from ethyl acetate or aqueous ethanol.

Found: C, 67.2, 66.9; H, 5.4, 5.7; N, 8.4, 8.5%.

Calculated for $C_{19}H_{20}O_4N_2$: C, 67.0; H, 5.9; N, 8.2%.

This substance was recovered unchanged from an attempted thiohydantoin reaction. Fractions (i), (ii), and (iii) above also consisted mainly of the α -anilide mixed with a small amount of the γ -isomer, the total yield of the latter being small. This is in contrast to the result with acetylglutamic anhydride, where the only product isolated was the γ -anilide.

(i) *Thiohydantoin from Phenylacetyl-DL-glutamic Acid γ -Anilide*

Application of the thiohydantoin reaction to the anilide above, having m.p. 159 °C., gave an oil, which was refluxed for 3 hours with 1N acetic acid. Evaporation and extraction of the residue with 2N ammonium bicarbonate to remove phenylacetic acid, gave a solid product, m.p. 210–212 °C., after recrystallization from aqueous ethanol, raised to 215 °C. on admixture with an authentic specimen of 2-thiohydantoyl-5- β -propionanilide, m.p. 216 °C. (Nicolet 1930).

(j) *1-Phenylacetyl-2-thiohydantoyl-5- β -propionamide*

The compound described by Baker and Jones (1951) as phenylacetyl-DL-glutamic acid α -amide was treated with ammonium thiocyanate and acetic anhydride in the usual way. Evaporation gave an oil, which, after extraction with a small volume of warm water, was recrystallized from aqueous ethanol. A 10% yield of the above 2-thiohydantoin was isolated having m.p. 177 °C.

Found: C, 55.3; H, 4.9%.

Calculated for $C_{14}H_{15}O_3N_3S$: C, 55.0; H, 5.0%.

Sodium fusion showed the presence of sulphur.

In this case, application of Le Quesne and Young's (1950) procedure gave no clear-cut separation of the two isomers. However, from one attempt to separate the mixed amides by crystallization, a small quantity of material was isolated, m.p. 133–134 °C. after two further recrystallizations from water. This material, thought originally (Swan 1952b) to be the true α -amide, is probably the γ -isomer, although the amount isolated was insufficient for direct characterization as the thiohydantoin.

Found: C, 59.4; H, 6.1; N, 10.5%.

Calculated for $C_{13}H_{16}O_4N_2$: C, 59.1; H, 6.1; N, 10.6%.

(k) *α - and γ -Hydrazides of Phenylacetyl-DL-glutamic Acid*

The method given by Baker and Jones (1951) for the α -isomer was followed. Using 3 g. of phenylacetyl-DL-glutamic anhydride, the first crop of crystals weighed 1.5 g., m.p. 177–179 °C. with evolution of gas. After recrystallization from water the product weighed 1.0 g. and had m.p. 183–184 °C., the melt remaining undecomposed, except for slight darkening, even when heated to 240 °C. Baker and Jones report m.p. 181 °C. (decomp.) for their product, which has proved to be the α -isomer as claimed (see below).

Found: N, 14.6%.

Calculated for $C_{13}H_{17}O_4N_3$: N, 15.0%.

Evaporation of the original mother liquor to about half-volume, and cooling for 48 hours, gave a large crop of hydrazine hydrochloride. This was filtered, the pH further adjusted until just acid to methyl orange, and the solution evaporated to dryness. Recrystallization of the residue from boiling water (25 ml.) gave 2.2 g. of crystals, melting at 143–145 °C. to a turbid liquid which decomposed with effervescence at 160–165 °C. After recrystallization from water using charcoal, the substance melted without decomposition at 145 °C. to a clear liquid, and on

further crystallization the m.p. rose to 147 °C. A mixture with the α -isomer melted at 140–145 °C., and analysis indicated that this compound was *phenylacetyl-DL-glutamic acid* γ -hydrazide.

Found: C, 55.9; H, 6.1; N, 14.3%.

Calculated for $C_{13}H_{17}O_4N_3$: C, 55.9; H, 6.1; N, 15.0%.

After removal of the α -derivative, the γ -hydrazide could also be isolated by evaporating to dryness, and extracting the residue with dioxane, in which hydrazine hydrochloride is sparingly soluble. Attempts to confirm the structure of the γ -isomer by conversion to a thiohydantoin were in this case unsuccessful.

(l) *Experiments using Phenylacetyl-DL-glutamic Acid* α -Hydrazide

(i) The compound above, m.p. 183–184 °C. (0.6 g.) was dissolved in excess 0.1N hydrochloric acid (4.4 ml.) and sodium nitrite (0.15 g.) added slowly, the temperature being kept at –5 to 0 °C. A large excess of aniline (1 ml.) was then added, together with sodium carbonate (0.35 g.), and the mixture acidified after standing for 15 minutes. Extraction with ether gave 0.6 g. of phenylacetyl-DL-glutamic acid α -anilide, m.p. 195 °C. showing no melting point depression with the substance described above.

(ii) The α -hydrazide (0.5 g.) was heated almost to boiling with a mixture of acetic anhydride (4.5 ml.), acetic acid (0.5 ml.), and ammonium thiocyanate. Successive evaporations with small amounts of water, and crystallization of the residue from a very small amount of ethanol gave a compound $C_{15}H_{17}O_4N_3$ in low yield (0.07 g.), m.p. 225–227 °C., raised to 228 °C. on recrystallization from 50% ethanol.

Found: C, 59.4; H, 5.5; N, 13.9%.

Calculated for $C_{15}H_{17}O_4N_3$: C, 59.4; H, 5.7; N, 13.9%.

Sodium fusion showed sulphur to be absent. The structure of this substance has not been further investigated.

(m) α - and γ -Amides of *p*-Nitrobenzoyl-DL-glutamic Acid

p-Nitrobenzoyl-DL-glutamic anhydride (3 g.) was allowed to react with liquid ammonia as described by King and Spensley (1950). The product separated as an oil after acidification and was taken into ethyl acetate. This solution was then extracted 10 times with 10 ml. portions of a solution of sodium carbonate (0.6 g.) in water (100 ml.). Each extract was acidified with concentrated hydrochloric acid (0.12 ml.) and allowed to stand overnight. In the first three fractions crystallization was very slow, in the others, separation of crystals occurred almost immediately. The weights of product and the melting points in each fraction were as follows:

(i) 0.16 g., 148–151 °C., (ii) 0.23 g., 187–190 °C., (iii) 0.20 g., 180 °C., (iv) 0.19 g., 178–181 °C. (a second crop, 0.02 g., m.p. 187–194 °C.), (v) 0.16 g., 175–178 °C., (vi) 0.12 g., 186–187 °C., (vii) 0.17 g., 190 °C., (viii) 0.11 g., 183–190 °C., (ix) 0.05 g., 175–178 °C., (x) 0.02 g., 210–214 °C.

Fraction (ii) was recrystallized from ethyl acetate and separated in microprisms, m.p. 195–196 °C. This is evidently *p*-nitrobenzoyl-DL-glutamine (King and Spensley (1950) report m.p. 191–193 °C.) which would be expected to be the stronger acid.

Found: C, 48.4; H, 4.2; N, 13.9%.

Calculated for $C_{13}H_{13}O_6N_3$: C, 48.8; H, 4.4; N, 14.2%.

Fractions (vi), (vii), and (viii) were combined and crystallized from methanol. Rosettes of needles separated, m.p. 188–190 °C., not altered by further crystallization. This compound was recovered unchanged from an attempted thiohydantoin reaction, and is evidently *p*-nitrobenzoyl-DL-isoglutamine. King and Spensley give m.p. 185 °C.

Found: C, 48.8; H, 4.4; N, 14.3%.

Calculated for $C_{13}H_{13}O_6N_3$: C, 48.8; H, 4.4; N, 14.2%.

Attempts to prepare pure materials from the other fractions were not successful. The second crop in fraction (ix) was shown by mixed melting point to be the γ -isomer (*p*-nitrobenzoyl-DL-glutamine). The main product in the above reaction is evidently the α -amide.

(n) Confirmation of the Structure of p-Nitrobenzoyl-DL-glutamic Acid α -Benzyl Ester

The ester (0.3 g.), prepared as described by King and Spensley (1950), was heated with acetic anhydride and ammonium thiocyanate in the usual manner. The product, 0.25 g., had m.p. 135–138 °C., raised to 146 °C. by crystallization from water containing a little ethanol, and showing no m.p. depression with the starting material.

(o) α - and γ -Anilides of p-Nitrobenzoyl-DL-glutamic Acid

p-Nitrobenzoyl-DL-glutamic anhydride (5 g.) in anhydrous dioxane (10 ml.) containing aniline (4 ml.) was allowed to stand overnight. Filtering and washing with dioxane gave 4.6 g. of product, m.p. 234–242 °C. (decomp.). Evaporation of the mother liquor gave a gum which solidified on shaking with 6N hydrochloric acid. Extraction into ethyl acetate and evaporation gave a second product, 2.1 g., which was purified by boiling with a small volume of ethyl acetate (30 ml.), cooling, and filtering. The insoluble material weighed 1.2 g. and melted at 199–200 °C. The main product above was sparingly soluble in most solvents but crystallized in long needles from nitrobenzene. Yield 3.9 g., m.p. 245 °C. This material was recovered unchanged after heating with acetic anhydride and ammonium thiocyanate and is therefore p-nitrobenzoyl-DL-glutamic acid α -anilide.

Found: C, 58.4; H, 4.8; N, 11.6%.

Calculated for $C_{18}H_{17}O_6N_3$: C, 58.2; H, 4.6; N, 11.3%.

The nitrobenzene mother liquor was extracted with sodium carbonate solution, acidification of which gave 0.1 g. of the lower melting isomer, m.p. 199–200 °C. This was combined with the 1.2 g. obtained from the reaction mother liquor, and a portion was purified for analysis by dissolving it in ammonium bicarbonate and precipitating with acid.

Found: C, 57.9; H, 4.8; N, 11.4%.

Calculated for $C_{18}H_{17}O_6N_3$: C, 58.2; H, 4.6; N, 11.3%.

p-Nitrobenzoyl-DL-glutamic acid γ -anilide has m.p. 199–200 °C., and was converted in the usual manner to 1-p-nitrobenzoyl-2-thiohydantoyl-5- β -propionanilide, m.p. 199–200 °C., depressed to 180–185 °C. on admixture with starting material.

Found: C, 55.3; H, 4.0; S, 7.6%.

Calculated for $C_{19}H_{16}O_5N_4S$: C, 55.3; H, 4.0; S, 7.9%.

IV. REFERENCES

- BAKER, W., and JONES, P. G. (1951).—*J. Chem. Soc.* **1951**: 1143.
 BARKER, C. C. (1951).—*Nature* **168**: 908.
 BERGMANN, M., STERN, F., and WITTE, C. (1926).—*Liebigs Ann.* **449**: 277.
 BERGMANN, M., and ZERVAS, L. (1928).—*Biochem. Z.* **203**: 280.
 HARINGTON, C. R., and OVERHOFF, J. (1933).—*Biochem. J.* **27**: 338.
 JOHNSON, T. B. (1913).—*Amer. Chem. J.* **49**: 68.
 JOHNSON, T. B., and GUEST, H. H. (1912a).—*Amer. Chem. J.* **47**: 242.
 JOHNSON, T. B., and GUEST, H. H. (1912b).—*Amer. Chem. J.* **48**: 103.
 JOHNSON, T. B., and NICOLET, B. H. (1913).—*Amer. Chem. J.* **49**: 197.
 KING, F. E., and SPENSLEY, P. C. (1950).—*J. Chem. Soc.* **1950**: 3159.
 LE QUESNE, W. J., and YOUNG, G. T. (1950).—*J. Chem. Soc.* **1950**: 1954.
 NICOLET, B. H. (1930).—*J. Amer. Chem. Soc.* **52**: 1192.
 SWAN, J. M. (1952a).—*Aust. J. Sci. Res. A* **5**: 711.
 SWAN, J. M. (1952b).—*Nature* **169**: 826.
 SYNGE, R. L. M. (1939).—*Biochem. J.* **33**: 1913.
 THIERFELDER, H. (1921).—*Hoppe-Seyl. Z.* **114**: 192.

THIOHYDANTOINS

III. ANHYDRIDE INTERMEDIATES IN THE FORMATION OF 1-ACYL-2-THIOHYDANTOINS FROM ACYLAMINO ACIDS

By J. M. SWAN*

[Manuscript received April 15, 1952]

Summary

Linear anhydrides are formed by the action of acetic anhydride on *p*-toluenesulphonylglycine and carbobenzyloxy-glycine, -phenylalanine, and - β -alanine. The anhydrides from the first two acids yield the corresponding 2-thiohydantoin with ammonium thiocyanate in acetic acid. Other methods for the preparation of 1-*p*-toluenesulphonylglycine anhydride, and its 2-thiohydantoin, are also given.

Hippuric acid, with ethyl chlorocarbonate and triethylamine yields ethyl hippurate, probably via the mixed anhydride and 2-phenyloxazol-5-one. The inclusion of ammonium thiocyanate gives 1-benzoyl-2-thiohydantoin in high yield. The *N*-carboxy anhydride of phenylalanine (4-benzyloxazolid-2,5-dione) also reacts with ammonium thiocyanate to give carbon dioxide and 5-benzyl-2-thiohydantoin.

I. INTRODUCTION

Johnson and Scott (1913) postulated that the intermediate in the formation of a 1-acyl-2-thiohydantoin from an acylamino acid was an oxazolone (see Part I of this Series, Swan 1952). However, the reaction succeeds with pyrrolidone-2-carboxylic acid, which cannot form an oxazolone, and Schlack and Kumpf (1926) therefore suggested that the initial intermediate in such cases was the linear acylaminoacetic mixed anhydride. Certain symmetrical linear anhydrides have now been converted to 2-thiohydantoins, and other methods for the preparation of these compounds discovered.

II. DISCUSSION

The action of hot acetic anhydride on carbobenzyloxyglycine gives the symmetrical *carbobenzyloxyglycine anhydride* which with aniline, and with glycine, gives the corresponding amide and a second mole of carbobenzyl-oxyglycine. With ammonium thiocyanate in acetic acid the anhydride gives 1-*carbobenzyloxy-2-thiohydantoin*, also prepared directly from carbobenzyl-oxyglycine in the usual way.

The anhydrides from carbobenzyloxy- β -alanine, carbobenzyloxyphenylalanine (not obtained pure), and *p*-toluenesulphonylglycine were characterized similarly by reaction with aniline. The last named anhydride is best prepared by the reaction of *p*-toluenesulphonylglycyl chloride with the silver salt of *p*-toluenesulphonylglycine, or by heating the acid with phosgene in toluene. 1-*p*-Toluenesulphonyl-2-thiohydantoin is formed from the acid in the usual way.

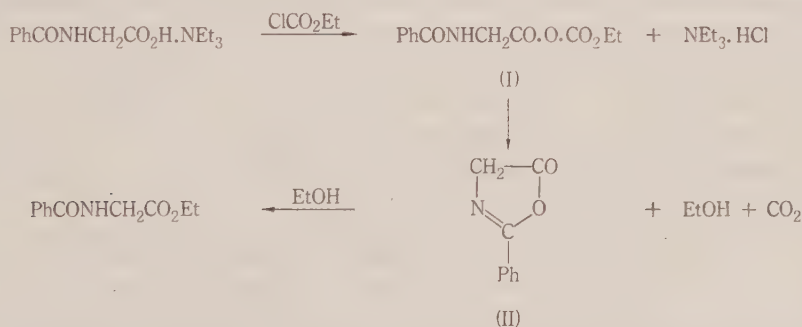
* Biochemistry Unit, Wool Textile Research Laboratory, C.S.I.R.O., Melbourne.

or from the acid chloride with lead or ammonium thiocyanate, or by including ammonium thiocyanate in the phosgene reaction. While this work was in progress Wieland, Kern, and Sehring (1950) reported the preparation of linear anhydrides of carbobenzyloxyamino acids by reaction of various salts of carbobenzyloxyamino acids with appropriate acid chlorides, and described the use of these anhydrides in the synthesis of peptides. This aspect of the present work was therefore discontinued.

Boissonnas (1951) has shown that ethyl chlorocarbonate can be used for the synthesis of carbobenzyloxy mixed anhydrides; the compounds $\text{PhCH}_2\text{OCONHCHRCO.O.CO}_2\text{Et}$ react readily with amino esters with liberation of carbon dioxide and ethanol, and formation of carbobenzyloxy peptide esters. Wieland and Bernhard (1951) have since made similar claims, while Vaughan (1951) has independently advocated the use of *sec*- or *isobutyl* chlorocarbonate for the same purpose. Linear mixed anhydrides between organic acids and simple acylamino acids such as aceturic acid, hippuric acid, etc. have not been described, although it has been suggested that such compounds may be intermediates in the formation of certain oxazol-5-ones (Clarke, Johnson, and Robinson 1949, p. 733).

In view of the ease with which oxazolones are converted to 2-thiohydantoins, the reaction between hippuric acid and ethyl chlorocarbonate has been studied.

The triethylamine salt of hippuric acid reacts vigorously with ethyl chlorocarbonate in dioxane, with separation of triethylamine hydrochloride. There is some evolution of carbon dioxide at this stage, and this continues over a period of several hours if the solution is allowed to stand. Under these circumstances the product isolated is ethyl hippurate.



The reaction appears to occur via the compounds I and II, since in the presence of benzaldehyde at 0 °C., an 11 per cent. yield of the well-known 2-phenyl-4-benzylideneoxazol-5-one (α -benzamido cinnamic azlactone) was obtained, together with ethyl hippurate and hippuric acid. Benzaldehyde is known to condense very readily with 2-phenyloxazol-5-one (II) (Clarke, Johnson, and Robinson 1949, p. 732). When the mixture of hippuric acid, triethylamine, and ethyl chlorocarbonate was heated slowly in dioxane to 50 °C., and furfural then added, only a trace of 2-phenyl-4-furfurylideneoxazol-5-one was obtained, together with a 48 per cent. yield of ethylhippurate. It does not seem likely

that aromatic aldehydes would condense directly with I, but the possibility that some, at least, of the ester is formed from I by loss of carbon dioxide without intervention of the cyclic intermediate II, or from I by an initial disproportionation to a symmetrical anhydride, cannot be excluded.

When ammonium thiocyanate is added to the mixture of hippuric acid, triethylamine, ethyl chlorocarbonate, and dioxane, 1-benzoyl-2-thiohydantoin is formed as the main product. No reaction occurs in the absence of the tertiary base. This technique offers an alternative procedure, in place of heating with acetic anhydride and ammonium thiocyanate, for the conversion of C-terminal residues in peptides and proteins to thiohydantoin.

It was thought that acetyl*is*thiocyanate might function both as a dehydrating agent and a source of thiocyanic acid, but no reaction was observed between hippuric acid and acetyl*is*thiocyanate at room temperature. However, the *is*thiocyanate can function as the source of thiocyanic acid in acetic anhydride, since on warming a mixture of hippuric acid, acetic anhydride, acetyl*is*thiocyanate, and benzene, 1-benzoyl-2-thiohydantoin is formed. While this paper was in course of preparation, Waley and Watson (1951) described the synthesis of 1-carbobenzoyloxyglycyl-5-methyl-2-thiohydantoin from carbobenzoyloxyglycylalanine and benzoyl*is*thiocyanate, the mixture being heated at 120 °C.

Finally, it has been found that 4-benzoyloxazolid-2,4-dione, the internal anhydride of *N*-carboxyphenylalanine (II; $R = C_6H_5CH_2$, $R = OH$), will also react with ammonium thiocyanate in acetic acid to give carbon dioxide and 5-benzyl-2-thiohydantoin. This reaction suggests itself as a convenient method for characterizing the oxazolid-2,5-diones, which have recently become important in the preparation of peptides (see, for example, Woodward and Schramm 1947; Astbury and Dagliesh 1948; Ambrose and Hamby 1949; Brown and Coleman 1949; Bailey 1950; Farthing 1950; Wessely, Riedl, and Tuppy 1950 etc.).

Other derivatives of acylamino acids, e.g. methyl aceturate, ethyl phenaceturate, phenylthiol phenaceturate, and *O*-hippuryl- α -benzaloxime (prepared from α -benzaloxime and 2-phenyloxazol-5-one) failed to give thiohydantoin with ammonium thiocyanate in acetic acid.

III. EXPERIMENTAL

(a) Carbobenzoyloxyglycine Anhydride

Carbobenzoyloxyglycine (1 g., m.p. 120 °C.) (Bergmann and Zervas 1932) was dissolved in a mixture of acetic anhydride (10 ml.) and ethyl acetate (10 ml.) and the solution was allowed to stand overnight. Evaporation under reduced pressure at 55 °C. afforded a crystalline solid which was filtered and washed with benzene, 1 g., m.p. 121 °C., mixed m.p., with carbobenzoyloxyglycine, 106 °C. The anhydride was recrystallized from chloroform, and separated in fine needles, m.p. 121 °C. Wieland, Kern, and Sehring (1950) give m.p. 118 °C.

Found: N, 7.0%.

Calculated for $C_{20}H_{20}O_7N_2$: N, 7.0%.

(b) Carbobenzoyloxyglycine Anilide

(i) The above anhydride (1.0 g.) was treated with aniline (0.9 ml.) when an exothermic reaction occurred. Addition of 2N hydrochloric acid precipitated a mixture of carbobenzyl-

oxyglycine (separated by its solubility in sodium bicarbonate) and its *anilide*, which crystallized from ethanol in rosettes of needles, 0.4 g., m.p. 145 °C.

Found: C, 67.6; H, 5.4; N, 10.3%.

Calculated for $C_{16}H_{16}O_3N_2$: C, 67.6; H, 5.7; N, 9.9%.

(ii) This substance was also prepared by reaction of carbobenzyloxyglycyl chloride (from 0.6 g. carbobenzyloxyglycine, Bergmann and Zervas 1932) and aniline at 0 °C. Yield 0.2 g., m.p. 145 °C.

(c) *Carbobenzyloxyglycylglycine*

Carbobenzyloxyglycine anhydride (1.1 g.) in ether (15 ml.) was shaken with glycine (1 g.) in 1N sodium hydroxide (13 ml.) for 7 hours and the aqueous layer then acidified. The precipitate weighed 0.5 g. and was separated by fractional crystallization from methanol into carbobenzyloxyglycylglycine, 0.2 g., m.p. 179 °C., showing no melting point depression with authentic material, and carbobenzyloxyglycine, 0.3 g., m.p. 120 °C. Using the anhydride (2.7 g.), ether (50 ml.), and glycine (4 g.) in water (2 ml.) containing sodium hydroxide (2.1 g.), the yield of mixed products was 3 g., from which approximately 0.8 g. of carbobenzyloxyglycylglycine was isolated.

(d) *1-Carbobenzyloxy-2-thiohydantoin*

(i) Carbobenzyloxyglycine (0.5 g.), acetic anhydride (2.0 ml.), acetic acid (0.2 ml.), and ammonium thiocyanate (0.22 g.) were heated at 55 °C. for 15–30 minutes. The solvents were then evaporated under reduced pressure, the oily residue shaken thoroughly with water, and the solid filtered and recrystallized from water containing a little ethanol. Yield of shining platelets, 0.3 g., m.p. 185–186 °C.

Found: C, 52.8; H, 4.5; N, 11.2%.

Calculated for $C_{11}H_{10}O_3N_2S$: C, 52.8; H, 4.0; N, 11.2%.

(ii) Carbobenzyloxyglycine anhydride (0.5 g.), acetic acid (0.5 ml.), and ammonium thiocyanate (0.2 g.) were heated together at 70 °C. for 5 minutes. Addition of water (10 ml.) precipitated an oil which soon crystallized, 0.18 g., m.p. 163–165 °C. The thiohydantoin crystallized from 40% aqueous ethanol in glistening plates, m.p. 185–186 °C., undepressed by admixture with the material above.

(e) *Carbobenzyloxyphenylalanine Anilide*

Carbobenzyloxyphenylalanine (2 g.) (Bergmann and Zervas 1932) was dissolved in a mixture of acetic anhydride (10 ml.) and dioxane (10 ml.) and the solution allowed to stand overnight. Evaporation under reduced pressure at 90 °C. gave a very viscous syrup. This material reacted slowly with aqueous alkali to regenerate carbobenzyloxyphenylalanine; and with aniline to give *carbobenzyloxyphenylalanine anilide* in low yield. Crystals from ethanol, m.p. 158–159 °C.

Found: C, 73.7; H, 5.9; N, 8.0%.

Calculated for $C_{23}H_{22}O_3N_2$: C, 73.8; H, 5.9; N, 7.5%.

(f) *Carbobenzyloxy-β-alanine Anhydride*

Carbobenzyloxy-β-alanine (2 g.) (Siffert and Du Vigneaud 1935) was dissolved in a mixture of dioxane (10 ml.) and acetic anhydride (10 ml.). Evaporation to dryness under reduced pressure afforded a crystalline solid which was filtered and washed with ether. Yield 1.1 g., m.p. 91 °C. The *anhydride* was recrystallized from a small volume of ethanol in clusters of glistening plates, m.p. 94 °C.

Found: C, 61.7; H, 5.6; N, 6.8%.

Calculated for $C_{22}H_{24}O_7N_2$: C, 61.7; H, 5.6; N, 6.5%.

(g) *Carbobenzyloxy-β-alanine Anilide*

The above symmetrical anhydride (0.3 g.) was heated at 100 °C. with aniline (0.3 ml.) for 15 minutes. Addition of 2N hydrochloric acid precipitated a white solid (0.3 g.) which was extracted with sodium bicarbonate solution to remove carbobenzyloxy-β-alanine. The insoluble product was recrystallized from aqueous ethanol, and then from benzene, m.p. 137 °C.

Found: C, 68.8; H, 5.9; N, 9.3%.

Calculated for $C_{17}H_{18}O_3N_2$: C, 68.5; H, 6.1; N, 9.4%.

The substance was soluble in ethanol, chloroform, and hot benzene, insoluble in cold water, cold benzene, and petrol (40–70 °C.).

All attempts to react the anhydride with ammonium thiocyanate to give a 6-membered analogue of a 2-thiohydantoin were unsuccessful, carbobenzyloxy-β-alanine generally being recovered in high yield.

(h) *p-Toluenesulphonylglycine Anilide*

(i) *p*-Toluenesulphonylglycine (0.5 g.) (Fischer and Bergmann 1913) was shaken with phosphorus pentachloride (0.6 g.) in dry ether at 0 °C. to give a clear solution. Evaporation gave *p*-toluenesulphonylglycyl chloride (0.5 g.) which crystallized from benzene in colourless needles, m.p. 82–83 °C., on addition of petrol (40–70 °C.).

Found: Cl, 14.0%.

Calculated for $C_9H_9O_3NSCl$: Cl, 14.4%.

This chloride (0.5 g.) was treated with aniline (0.8 g.) in ether, the mixture shaken with 2N hydrochloric acid, and the insoluble product filtered, 0.6 g. The *anilide* crystallized from ethanol in glistening needles, m.p. 159 °C.

Found: C, 59.9; H, 5.4; N, 9.4%.

Calculated for $C_{15}H_{16}O_3N_2S$: C, 59.2; H, 5.3; N, 9.2%.

(ii) *p*-Toluenesulphonylglycine (1.05 g.) was dissolved in warm acetic anhydride (10 ml.) and the solution then evaporated to dryness at 100 °C. After evaporating twice more with toluene the residual oil weighed 1 g. On addition of aniline heat was evolved, and a solid separated. The mixture was treated with 2N hydrochloric acid, the product filtered, and crystallized from aqueous ethanol to give needles, 0.23 g., m.p. 159 °C., undepressed by admixture with the product above.

(i) *p-Toluenesulphonylglycine Anhydride*

(i) *p*-Toluenesulphonylglycine (2.3 g.) was refluxed for 2 hours with a 12.5% solution of phosgene in toluene (42 ml.) diluted with dry toluene (50 ml.); a further 42 ml. of the phosgene solution was added and refluxing was continued for 2 hours. From the crystalline mixture which separated on cooling, some unchanged starting material was extracted with boiling chloroform, the insoluble material being *p*-toluenesulphonylglycine *anhydride*, 0.7 g. This was crystallized several times by dissolving in a very small volume of dry acetone and precipitating with benzene, clusters of felted needles, m.p. 167–168 °C.

Found: C, 48.8; H, 4.7; S, 14.8%.

Calculated for $C_{18}H_{20}O_7N_2S_2$: C, 49.1; H, 4.6; S, 14.5%.

(ii) *p*-Toluenesulphonylglycyl chloride (2.2 g.) was shaken in chloroform for 3 hours with the silver salt of *p*-toluenesulphonylglycine. The mixture was then filtered, the solvent evaporated, and the residual gum dissolved in acetone (4 ml.). Addition of benzene (20 ml.) gave a solid (2.0 g.) which was filtered and recrystallized first from chloroform-benzene, and then from a small volume of chloroform by the addition of petrol (40–60 °C.). The anhydride was identified by m.p. 167–168 °C. and by its conversion to *p*-toluenesulphonylglycine *anilide*, m.p. 159 °C.

(j) Aniline Salt of p-Toluenesulphonylglycine

This salt separated on the addition of aniline to a benzene solution of the acid, and crystallized from water in small glistening plates, m.p. 133–135 °C.

Found: C, 56.1; H, 5.6; S, 10.0, 9.8%.

Calculated for $C_{15}H_{18}O_4N_2S$: C, 55.9; H, 5.6; S, 9.9%.

An aqueous solution of the salt gave positive tests for aniline.

(k) 1-p-Toluenesulphonyl-2-thiohydantoin

(i) *p*-Toluenesulphonylglycine (0.5 g.), acetic anhydride (4.5 ml.), acetic acid (0.5 ml.), and ammonium thiocyanate (0.2 g.) were heated at 70 °C. for 30 minutes, and the mixture worked up in the usual way. On crystallization from 50% aqueous ethanol, a very small yield (30 mg.) of 1-*p*-toluenesulphonyl-2-thiohydantoin was obtained. Short needles, m.p. 235 °C. (decomp.), after recrystallization.

Found: C, 44.9; H, 3.6; N, 10.5%.

Calculated for $C_{10}H_{10}O_3N_2S_2$: C, 44.4; H, 4.0; N, 10.4%.

(ii) *p*-Toluenesulphonylglycyl chloride (0.17 g.) was shaken overnight with lead thiocyanate (0.15 g.) in benzene and the mixture was then heated to boiling, cooled, and filtered. Evaporation of the filtrate and crystallization of the residue from aqueous ethanol gave a small yield of colourless flat needles of the thiohydantoin, m.p. 235 °C. (decomp.).

Found: N, 10.4%.

(iii) *p*-Toluenesulphonylglycyl chloride (0.7 g.) was warmed with ammonium thiocyanate (1 g.) in acetic acid (5 ml.). A deep red colour appeared first which soon changed to pale yellow, and a cloudy precipitate appeared. Cooling gave the thiohydantoin as a thick crystalline precipitate which was filtered after the addition of water. Yield 0.5 g., 65.5%, m.p. 235 °C. (decomp.).

(iv) A mixture of *p*-toluenesulphonylglycine (0.87 g.), a 12.5% solution of phosgene in toluene (18 ml.), ammonium thiocyanate (0.5 g.), and acetic acid (3 ml.) was refluxed for 1.5 hours. After decantation from the sticky insoluble solid, the solvents were evaporated and the residue recrystallized twice from aqueous ethanol. Yield of the 2-thiohydantoin was 0.2 g., m.p. 235 °C. (decomp.).

(l) 1-Benzoyl-2-thiohydantoin

Hippuric acid (1.0 g.), acetic anhydride (9 ml.), acetic acid (1.0 ml.), and 5 ml. of a solution of acetylthiocyanate in benzene (prepared by reacting lead thiocyanate (3 g.) with acetyl chloride (1.25 ml.) in benzene) were heated on the steam-bath for 5 minutes, the solution was allowed to stand for 1 hour, and the solvents then evaporated under reduced pressure. Crystallization of the residue from aqueous ethanol with charcoal gave flat needles, m.p. 165–168 °C., unchanged by admixture with an authentic specimen of 1-benzoyl-2-thiohydantoin.

(m) Experiments using Ethyl Chlorocarbonate

(i) Hippuric acid (3.6 g.), triethylamine (2.2 ml.), and dioxane (50 ml.) were warmed to give a clear solution and cooled to 0 °C. Ethyl chlorocarbonate (2.2 ml.) was added, and the temperature allowed to rise slowly to room temperature. There was a steady evolution of carbon dioxide during this period, which continued as the temperature was raised slowly to the boiling point. The total volume of gas collected was 360 ml. The solution was then cooled, the triethylamine hydrochloride filtered, and the solvent evaporated. The residual oil was dissolved in ether, extracted with ammonium bicarbonate to remove hippuric acid, and the ether evaporated to give ethyl hippurate, weighing 2.2 g. after recrystallization from water containing a little ethanol. The ester had m.p. 60 °C., undepressed by admixture with an authentic sample. The hippuric acid recovered weighed 1.0 g., m.p. 186–187 °C.

(ii) The above experiment was repeated, but with the inclusion of benzaldehyde (2 ml.). The reactants were kept at 0 °C. for 24 hours, then at room temperature for 24 hours, the solvent being then evaporated under reduced pressure. Extraction with 2N sodium carbonate of an ethyl acetate solution of the residual oil caused crystallization of 2-phenyl-4-benzylideneoxazol-5-one, 0.6 g., m.p. 164–165 °C., undepressed by admixture with an authentic specimen. Acidification of the sodium carbonate extract gave hippuric acid (1.0 g.) while evaporation of the ethyl acetate solution and crystallization of the residue gave ethyl hippurate (1.3 g.).

When the reactants were first heated to 50 °C. and furfural (2 ml.) added, together with a few drops of pyridine, a trace only of 2-phenyl-4-furfurylideneoxazol-5-one, m.p. 171 °C., was isolated, together with 2 g. of ethyl hippurate.

(iii) Hippuric acid (1.8 g.) was dissolved in dioxane (30 ml.) by the addition of triethylamine (1.1 ml.) and the mixture was treated with ethyl chlorocarbonate (1.1 ml.). Ammonium thiocyanate (1.0 g.) was added, the solution heated to boiling, and allowed to stand overnight. Evaporation and crystallization from aqueous ethanol gave 1-benzoyl-2-thiohydantoin, 1.4 g., m.p. 166–167 °C. When the ammonium thiocyanate was added in solution in cold acetic acid (5 ml.) and the mixture allowed to stand for 12 hours, the product weighed 0.8 g., m.p. 165–167 °C.

(iv) Hippuric acid (1.0 g.), dioxane (10 ml.), and ammonium thiocyanate (0.5 g.) were heated to give a clear solution and the mixture cooled. On addition of ethyl chlorocarbonate (0.6 ml.) no apparent reaction occurred, but on further addition of triethylamine (0.6 ml.) heat was evolved and a precipitate appeared. After standing for 18 hours the dioxane was evaporated, the residue digested with cold water containing a trace of acetone, and the product filtered, 0.71 g., m.p. 166–167 °C. after recrystallization. Partial evaporation of the initial aqueous filtrate gave hippuric acid, 0.21 g. The yield of thiohydantoin on hippuric acid consumed was 79%.

(n) 5-Benzyl-2-thiohydantoin

4-Benzylloxazol-2,5-dione (0.8 g.) (Farthing 1950; Levy 1950) was added to a cold solution of ammonium thiocyanate (0.5 g.) in acetic acid (10 ml.). On warming to 50 °C., carbon dioxide was evolved and the colour changed from red to yellow. Evaporation of the solvent and crystallization of the residue from aqueous ethanol gave 5-benzyl-2-thiohydantoin, m.p. 180–182 °C., undepressed by admixture with an authentic specimen.

(o) O-Hippuryl- α -benzaldehyde oxime

A mixture of 2-phenylloxazol-5-one (Clarke, Johnson, and Robinson 1949, p. 781), α -benzaldehyde oxime (0.7 ml.), benzene (5 ml.), and one drop of triethylamine, was allowed to stand for 16 hours. Colourless crystals separated, 1.0 g., m.p. 124–126 °C., and the product recrystallized from aqueous ethanol as needles, m.p. 135 °C.

Found: C, 68.1; H, 5.1%.

Calculated for $C_{18}H_{14}O_5N_2$: C, 68.1; H, 5.0%.

This compound was insoluble in dilute acid and alkali but dissolved slowly in warm 2N sodium hydroxide. On dissolving in chloroform saturated with dry ammonia, heat was evolved and hippuramide separated, m.p. (and mixed m.p. with a specimen prepared by treating 2-phenylloxazol-5-one with ammonia) 185 °C. The filtrate had a strong odour of benzaldehyde and treatment with 2,4-dinitrophenylhydrazine in 1N hydrochloric acid gave benzaldehyde 2,4-dinitrophenylhydrazone, m.p. 235 °C. O-Hippuryl- α -benzaldehyde oxime failed to give any crystalline product when heated with ammonium thiocyanate in acetic acid.

IV. REFERENCES

- AMBROSE, E. J., and HANBY, W. E. (1949).—*Nature* **163**: 483.
 ASTBURY, W. T., and DALGLIESH, C. E. (1948).—*Nature* **162**: 596.
 BAILEY, J. L. (1950).—*J. Chem. Soc.* **1950**: 3461.
 BERGMANN, M., and ZERVAS, L. (1932).—*Ber. deutsch. chem. Ges.* **65**: 1192.
 BOISSONNAS, R. A. (1951).—*Helv. Chim. Acta* **34**: 874.
 BROWN, C. J., and COLEMAN, D. (1949).—*Nature* **163**: 834.

- CLARKE, H. T., JOHNSON, J. R., and ROBINSON, R. (Editors) (1949).—"The Chemistry of Penicillin." (Princeton Univ. Press.)
- FARTHING, A. C. (1950).—*J. Chem. Soc.* **1950**: 3213.
- FISCHER, E., and BERGMANN, M. (1913).—*Liebigs Ann.* **398**: 117.
- JOHNSON, T. B., and SCOTT, W. M. (1913).—*J. Amer. Chem. Soc.* **35**: 1136.
- LEVY, A. L. (1950).—*Nature* **165**: 152.
- SCHLACK, P., and KUMPF, W. (1926).—*Hoppe-Seyl. Z.* **154**: 125.
- SIFFERD, R. H., and DU VIGNEAUD, V. (1935).—*J. Biol. Chem.* **108**: 753.
- SWAN, J. M. (1952).—*Aust. J. Sci. Res. A* **5**: 711.
- VAUGHAN, J. R. (1951).—*J. Amer. Chem. Soc.* **73**: 3547.
- WALEY, S. G., and WATSON, J. (1951).—*J. Chem. Soc.* **1951**: 2394.
- WESSELY, F., RIEDL, K., and TUPPY, H. (1950).—*Mh. Chem.* **81**: 861.
- WIELAND, T., and BERNHARD, H. (1951).—*Liebigs Ann.* **572**: 190.
- WIELAND, T., KERN, W., and SEHRING, R. (1950).—*Liebigs Ann.* **569**: 117.
- WOODWARD, R. B., and SCHRAMM, C. H. (1947).—*J. Amer. Chem. Soc.* **69**: 1551.

THE REACTION OF AMINO ACIDS AND PROTEINS WITH DIAZONIUM COMPOUNDS

I. A SPECTROPHOTOMETRIC STUDY OF AZO-DERIVATIVES OF HISTIDINE AND TYROSINE

By H. G. HIGGINS* and DALLAS FRASER*

[Manuscript received April 1, 1952]

Summary

The coupling in alkaline solution of free histidine and tyrosine with *p*-diazobenzene-sulphonic acid has been studied as a basis for quantitative investigation of the behaviour of proteins in the presence of diazonium compounds. Both mono- and bisazo-derivatives may be formed with each amino acid, the wavelengths of maximum absorption being 380 and 490 m μ respectively for histidine and 325 and 480 m μ respectively for tyrosine; the wavelength displacements on accepting a second diazonium group are comparable to effects reported for similarly conjugated systems. When appropriate corrections are made for other absorbing components, the extinctions at 490 m μ and the initial histidine concentrations up to the (bis) equivalent of diazonium compound conform to Beer's law, showing that bis(azobenzenesulphonic acid)histidine is formed exclusively under these conditions. At higher histidine concentrations mono(azobenzenesulphonic acid)histidine is formed in increasing proportions. Even at low initial tyrosine concentrations both mono- and bis(azobenzenesulphonic acid)-tyrosine are formed, although mono-coupling is again favoured more as the concentration increases. No direct reaction between histidine and the nitrous acid used in the formation of the diazonium compound could be detected spectrophotometrically, but diazotization of tyrosine appears to reduce the propensity for bis-coupling to some extent. If urea is incorporated in the system to destroy excess nitrous acid, a yellow compound having maximum absorption at 430 m μ is formed between *p*-diazobenzenesulphonic acid and the urea.

I. INTRODUCTION

The coupling of diazonium salts with amino acids and proteins is of wide interest in protein chemistry, physiology, and immunology. By varying the nature of the radical attached, usually in the *p* position, to the benzene ring of the diazonium salt, Landsteiner (1915) and others have prepared many azo-protein compounds, some of which have special clinical and pharmacological properties. The reaction with histidine was described by Pauly (1905, 1915) and now forms the basis of various colorimetric methods of determining both histidine and histamine. Pauly (1915) also studied the coupling of diazonium salts with tyrosine and concluded from analysis of the sparingly soluble compound formed with *p* diazobenzenesulphonic acid that the tyrosine, like histidine, coupled with the diazonium salt in the ratio 1 : 2. The linkage, for histidine, is through the ring carbon atoms (Pauly 1905) and the reaction with tyrosine (Pauly 1915 ;

* Division of Forest Products, C.S.I.R.O., Melbourne.

Herriott 1947) takes place at the positions *ortho* to the hydroxyl group. The work of Boyd and Hooker (1934) and Boyd and Mover (1935) indicated that protein groups other than the imidazole and the phenolic might participate in diazonium coupling. Eagle and Vickers (1936) showed that amino- and imino-groups can react with diazonium salts but that only in the case of tryptophane were the resulting compounds more coloured than the diazonium salt itself. In his review Herriott suggests that spectrophotometric studies might prove valuable in identifying the protein groups concerned in the coupling reaction. In the visible spectrum such studies will be concerned mainly with the coupled products of histidine and tyrosine and perhaps tryptophane.

The present investigation arose out of a critical examination of the method used by MacPherson (1942) for estimating histidine, and it is concerned principally with the coupling on ring carbon atoms, which leads to the production of strongly coloured compounds. Of particular interest were the questions of distinguishing spectrophotometrically between the azo-derivatives of histidine and tyrosine and of applying similar methods to studying the occurrence, accessibility, and behaviour of these amino acids in intact protein as well as hydrolysates. In this respect the work of Albanese *et al.* (1949) has indicated that the method of Albanese, Frankston, and Irby (1945) for determining urinary histidine colorimetrically (which is also based on the Pauly reaction) is applicable to non-diffusible fractions of enzymatic casein digests and human urine.

The present paper describes the coupling of free amino acids; a later communication will be concerned with proteins and protein hydrolysates.

II. GENERAL PROCEDURE

Except where otherwise indicated, the method of MacPherson (1942), which is a development of that used by Hanke and Koessler (1920), Hanke (1925), and Jorpes (1932), was followed as a basis for the formation of the azo-compounds. The reagents used were:

Sulphanilic acid: 1 per cent. solution of anhydrous material in 10 per cent. HCl.

Sodium nitrite: 5 per cent. solution.

Sodium carbonate: 20 per cent. solution.

Ethanol: 75 per cent. solution containing 1 ml. 20 per cent. Na_2CO_3 per 100 ml.

Histidine: B.D.H. L(+)-histidine hydrochloride, found by Volhard method to contain 0.91 equivalents of chloride per mole of histidine.

Tyrosine: B.D.H. L-tyrosine.

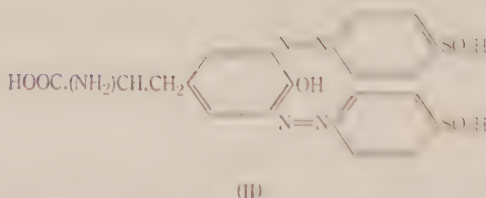
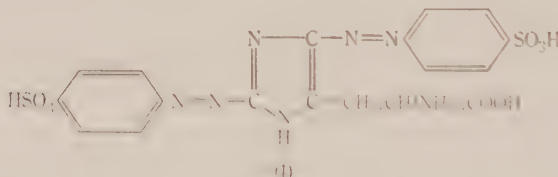
An aliquot of the amino acid solution not exceeding 10 ml. is transferred to a 25 ml. flask, and 1 ml. sulphanilic acid is added, followed by 1 ml. sodium nitrite. After 30 minutes at room temperature, 3 ml. sodium carbonate is added and, after mixing, 10 ml. ethanol. The solution is made up to volume with distilled water. For histidine, MacPherson recommends that the solution should contain 0.01–0.20 mg.; in the studies reported below a very much wider

range of histidine concentrations was used. Absorption spectra were determined by means of a Unicam S.P. 500 photoelectric quartz spectrophotometer, using 1 cm. cells, and water as the reference medium. After completion of the reaction and making up to volume, the coloured solutions were diluted with water as required to permit accurate measurement of extinctions, and appropriate corrections were made subsequently.

III. THEORETICAL CONSIDERATIONS

For the reactants and products giving rise to light absorption, the initial components of the system may be regarded as sulphanilic acid, nitrous acid, and the amino acid (histidine or tyrosine). The essential reactions are then :

- (i) The diazotization of the sulphanilic acid by nitrous acid to produce *p*-diazobenzenesulphonic acid (an equimolar reaction).
- (ii) The coupling of this diazonium compound with the amino acid under alkaline conditions, in which, according to Pauly (1915), one mole of histidine or tyrosine reacts with 2 moles of the diazonium compound to produce 1 mole of bis(azobenzenesulphonic acid)histidine (I) or bis(azobenzenesulphonic acid)tyrosine (II).



In addition to these reactions the possibility exists of the amino acid reacting directly with nitrous acid. The work of Wiley and Lewis (1930) suggested that histidine might be involved in this way, but it will be shown later that the absorption given by products of this reaction can be neglected against that given by the azo-derivatives. Diazotization of tyrosine, studied by Philpot and Small (1938*a*), is, however, a complicating factor, and will be discussed later. The systems containing histidine are therefore more amenable to theoretical treatment.

In the system initially comprising histidine, sulphanilic acid, and nitrous acid, and subjected to a change from acid to alkaline conditions after the diazotization of the sulphanilic acid, the following notation is adopted :

With Appropriate Subscript

c = initial molar concentration of reactant (including diazonium salt upon formation),

C = final molar concentration of reactant or product,

c' , C' = corresponding concentrations in g./l.,

M = molecular weight,

ε = molecular extinction coefficient,

ε' = molar extinction coefficient,

E = extinction due to initial concentration of an individual reactant,

Without Subscript

E = final extinction of whole system.

Subscripts "diazo", "HONO", "sulph", "hist", "mono", and "bis" refer respectively to the diazonium salt, nitrous acid, sulphanilic acid, histidine, and mono- and bishistidine azo-derivatives. The spectrophotometric evidence reveals the existence of only one monoazo-derivative.

It appears that in reaction (i) above, the sulphanilic acid is converted to the diazonium compound in practically the theoretical amount under the conditions employed here. This is based on the observation that the concentration of bis(azobenzenesulphonic acid)histidine subsequently formed by coupling is proportional to the initial histidine concentration up to the bis equivalent of the sulphanilic acid (see Figs. 5 and 6 (b), and relevant discussion in Section IV (c)). Then, since $c_{\text{HONO}} > c_{\text{sulph}}$ (from the procedure adopted), we have

$$c_{\text{diazo}} = c_{\text{sulph}},$$

$$C_{\text{HONO}} = c_{\text{HONO}} - c_{\text{sulph}},$$

and

$$C_{\text{sulph}} = 0.$$

Histidine itself does not absorb to any appreciable extent in the visible or near ultraviolet, so that we may take

$$\varepsilon_{\text{hist}} = 0.$$

For the final state of the system and for a particular wavelength,

$$E = \sum C\varepsilon$$

$$= C_{\text{hist}}\varepsilon_{\text{hist}} + C_{\text{sulph}}\varepsilon_{\text{sulph}} + C_{\text{HONO}}\varepsilon_{\text{HONO}} + C_{\text{diazo}}\varepsilon_{\text{diazo}} + C_{\text{mono}}\varepsilon_{\text{mono}} + C_{\text{bis}}\varepsilon_{\text{bis}},$$

which reduces to

$$E = (c_{\text{HONO}} - c_{\text{sulph}}) \frac{E_{\text{HONO}}}{C_{\text{HONO}}} + C_{\text{diazo}} \frac{E_{\text{diazo}}}{c_{\text{sulph}}} + C_{\text{mono}}\varepsilon_{\text{mono}} + C_{\text{bis}}\varepsilon_{\text{bis}} \quad \dots \dots \dots (1)$$

In the absence of histidine ($c_{\text{hist}} = 0$),

$$C_{\text{mono}} = 0, \quad C_{\text{bis}} = 0, \quad C_{\text{diazo}} = c_{\text{diazo}} = c_{\text{sulph}}.$$

This may be regarded as the control state and, if E_{cont} now represents the final extinction of the system, we have from (1)

$$E_{\text{cont}} = \left(1 - \frac{c_{\text{sulph}}}{c_{\text{HONO}}}\right) E_{\text{HONO}} + E_{\text{diazo}} \quad \dots \dots \dots (2)$$

When $c_{\text{hist}} < \frac{1}{2}c_{\text{sulph}}$, the coupling reaction can be expected to proceed to the formation of the bisazo-compound, so that

$$C_{\text{diaz}} = c_{\text{sulph}} - 2c_{\text{hist}}$$

and

$$C_{\text{bis}} = c_{\text{hist}}.$$

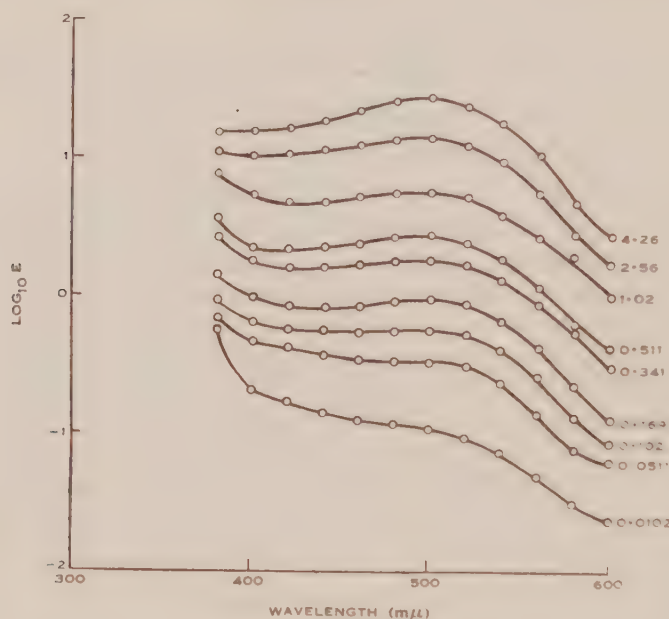


Fig. 1.—Uncorrected absorption spectra at lower histidine concentrations. Figures beside each curve are values of c' (mg. histidine/25 ml.).

Then, from (1)

$$E = (c_{\text{HONO}} - c_{\text{sulph}}) \frac{E_{\text{HONO}}}{c_{\text{HONO}}} + (c_{\text{sulph}} - 2c_{\text{hist}}) \frac{E_{\text{diaz}}}{c_{\text{sulph}}} + c_{\text{hist}} \epsilon_{\text{bis}}. \quad (3)$$

Since E_{diaz} is independent of c_{hist} , we may substitute for it from (2), whence, after rearrangement,

$$c_{\text{hist}} \epsilon_{\text{bis}} = E - \left(1 - \frac{c_{\text{sulph}}}{c_{\text{HONO}}}\right) E_{\text{HONO}} - \left(1 - 2 \frac{c_{\text{hist}}}{c_{\text{sulph}}}\right) \left(E_{\text{cont}} - \left(1 - \frac{c_{\text{sulph}}}{c_{\text{HONO}}}\right) E_{\text{HONO}}\right). \quad (4)$$

whence, since $c' = cM$,

$$c_{\text{hist}} \epsilon_{\text{bis}} = E - E_{\text{cont}} + 2 \frac{c'_{\text{hist}} M_{\text{sulph}}}{c'_{\text{sulph}} M_{\text{hist}}} E_{\text{cont}} - \left(1 - \frac{c'_{\text{sulph}} M_{\text{HONO}}}{c'_{\text{HONO}} M_{\text{sulph}}}\right) E_{\text{HONO}}. \quad (5)$$

Upon substitution of the appropriate molecular weights and initial concentrations of reactants used, this reduces to

$$c_{\text{hist}\varepsilon_{\text{bis}}} = E - E_{\text{cont}} + 0.223c''_{\text{hist}}(E_{\text{cont}} - 0.920E_{\text{HONO}}), \quad \dots\dots\dots (6)$$

where c''_{hist} is now the number of mg. in the system of volume 25 ml.

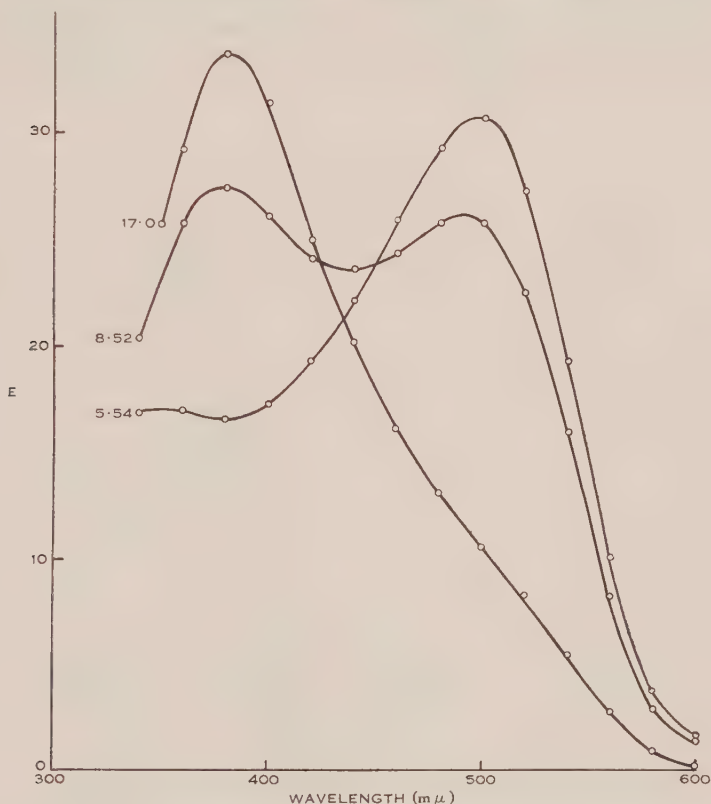


Fig. 2.—Absorption spectra at higher histidine concentrations ($c_{\text{hist}} > \frac{1}{2}c_{\text{sulph}}$), showing increase in mono- and decrease in bis-coupling as concentration rises. Figures beside each curve are values of c'' (mg. histidine/25 ml.).

IV. RESULTS AND DISCUSSION

(a) Absorption Spectra Observed with Histidine

Figures 1 and 2 show the observed absorption spectra in the visible region for molar concentrations of histidine less than and greater than half the molar concentration of sulphanilic acid respectively. The spectra of Figure 2 and the higher concentrations of Figure 1 represent a deep red colour, while, with decreasing concentration, the spectra of Figure 1 represent an increasing proportion of yellow. The red colour appears only on the addition of alkali to the system and the absorption maximum at *c.* 490 $m\mu$ is undoubtedly to be attributed to a histidine azo-compound. If the spectra are corrected according to equation

(6) they appear as shown in Figure 3, and even at very low histidine concentration the absorption band at 490 m μ is unmasked.

(b) Control Spectra

The values of E_{cont} and E_{HONO} used in correcting the observed values of E were taken from a set of control spectra measured with the object of isolating the absorption characteristics of individual components of the system. These spectra, which are shown in Figure 4, were all obtained by following MacPherson's procedure, but omitting certain constituents. For the nitrous acid (curve 2)

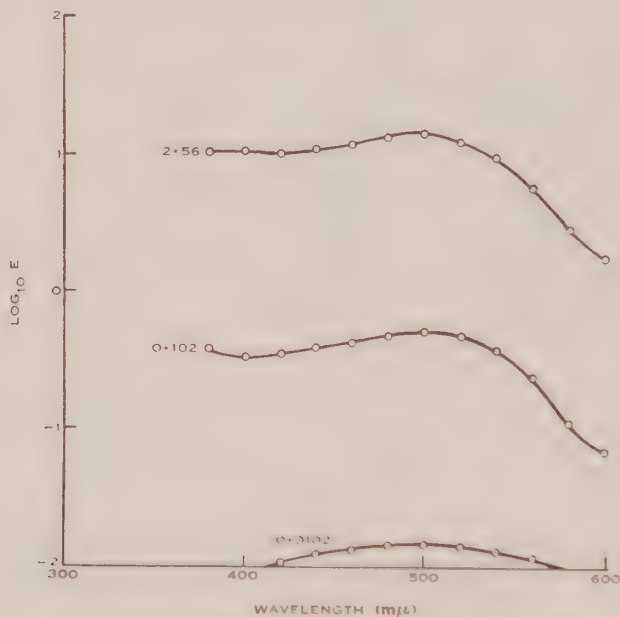


Fig. 3.—Absorption spectra corrected according to equation (6). Figures beside each curve are values of c'' (mg. histidine, 25 ml.).

a prominent absorption band due to ONO^- occurs at $c. 355 \text{ m}\mu$, as noted by Philpot and Small (1938*b*) in their study of the action of nitrous acid on pepsin. The maximum absorption of *p*-diazobenzenesulphonic acid is well into the ultraviolet, as shown by curve 1 of Figure 4, which is the resultant of the extinctions due to this substance and to unreacted nitrous acid. Sulphanilic acid does not absorb in the region above 300 m μ .

(c) Bis-coupling with Histidine

Pauly's conclusion, arrived at by analytical methods, that each mole of histidine can couple with two moles of diazonium salt, is given strong support by plotting the extinction due to the coupled compound against the histidine concentration. These relations are shown in Figures 5 and 6 for 480 and 500 m μ respectively; for the concentrations plotted in Figure 5 the portion of E not due to the azo-compound is negligible. Beer's law is followed until the molar

concentration of histidine reaches half that of the sulphanilic acid in the system, as indicated by the vertical broken line; at higher concentrations deviation from Beer's law is evident. At the lowest histidine concentrations excellent

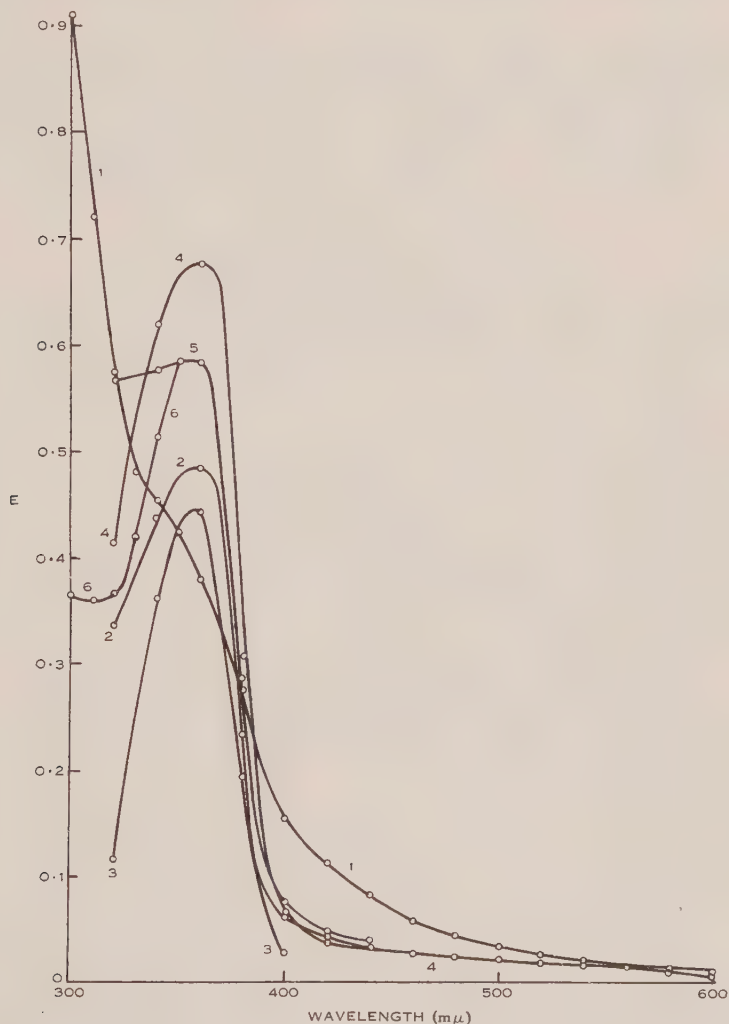


Fig. 4.—Control spectra. 1, Sulphanilic acid + sodium nitrite + sodium carbonate + ethanol ($E \times \frac{1}{2}$); 2, hydrochloric acid (without sulphanilic acid) + sodium nitrite + sodium carbonate + ethanol; 3, as for curve 2, but without ethanol; 4, as for curve 2, +1.02 mg. histidine/25 ml.; 5, as for curve 2, +5.96 mg. histidine/25 ml.; and 6, as for curve 2, +4.26 mg. histidine/25 ml.

concordance with Beer's law is observed upon the application of equation (6) (Fig. 6 (a)), whereas pronounced deviations are observed when the appropriate corrections are not applied to the measured extinctions. Thus MacPherson's contention that the colour produced follows Beer's law requires modification

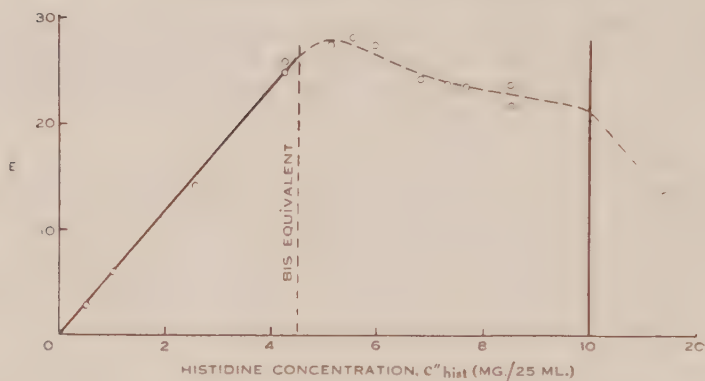


Fig. 5.—Dependence of final extinction E , at 480 mμ, upon initial histidine concentration.

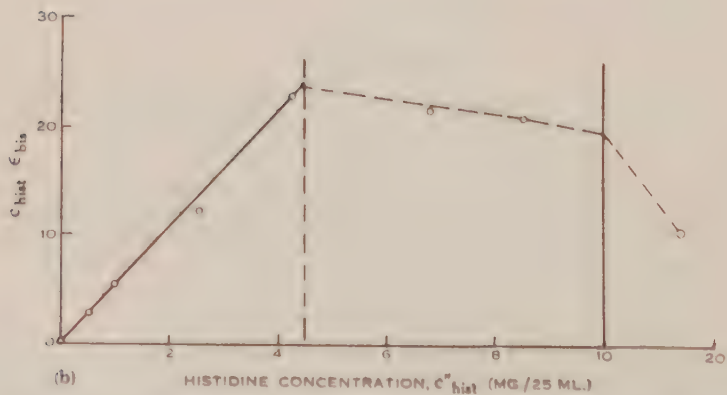
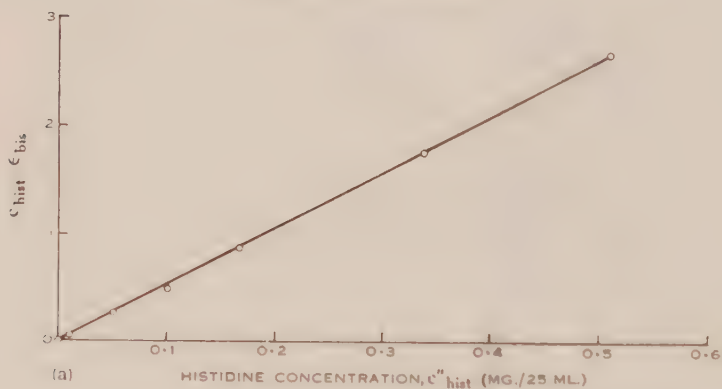


Fig. 6.—Dependence of final extinction due to azo-compound, at 500 mμ, upon initial histidine concentration, showing conformity with Beer's law up to the bis equivalent. (a) Low concentration. (b) Complete range.

in the light of these data, and it is further apparent that the amount of histidine which may be estimated at 480–500 $m\mu$ by MacPherson's method is very much greater than the upper limit he recommends, viz. 0.2 mg., and does in fact exceed 4 mg.

In so far as the assumption of coupling on the two carbon atoms of the imidazole ring leads to such good concordance between the predicted and observed relations between histidine concentration and the final extinction attributable to the azo-derivative, there can be little doubt that the absorption band at 490 $m\mu$ represents the bisazo-compound (I), and that this is formed exclusively, or practically so, when $c_{\text{hist}} \leq \frac{1}{2}c_{\text{sulph}}$.

(d) *Mono-coupling with Histidine*

At higher histidine concentrations, however, it is evident from Figures 5 and 6 (b) that the assumption of exclusive bis-coupling fails to account for the observed extinctions at 490 $m\mu$, which fall off with increasing histidine concentration. The most plausible interpretation of this behaviour lies in the ability of the imidazole ring to accept either one or two diazonium groups. Strong support for this view can be found in the appearance, when c_{hist} exceeds $\frac{1}{2}c_{\text{sulph}}$, of a pronounced absorption band at 380 $m\mu$ (Fig. 2), which can be attributed to mono(azobenzenesulphonic acid)histidine. The transformation of the monoazo-compound to the bisazo-compound increases the conjugation of the chromophoric system and an increase in the wavelength of maximum absorption of the order observed (from 380 to 490 $m\mu$) is quite in accordance with analogous effects observed in similar conjugated systems (e.g. Piper and Brode 1935).

(e) *Molar Extinction Coefficients of Histidine Azo-Derivatives*

The molar extinction coefficients at 380 and 490 $m\mu$, the wavelengths for maximum absorption for the mono- and bis-compounds respectively, may be calculated from observed extinctions. To determine ϵ_{mono} the histidine concentration was increased until the extinction at 380 $m\mu$ no longer increased appreciably with concentration. The limiting values of E , after correction for excess nitrous acid, were divided by c_{sulph} (representing the maximum histidine concentration which can participate in mono-coupling) to give ϵ_{mono} . At lower histidine concentrations ($c_{\text{hist}} < \frac{1}{2}c_{\text{sulph}}$), where coupling is exclusively of the bis-type, ϵ_{bis} may be calculated from equation (6) or from the slopes of Figures 5 and 6. The molar extinction coefficients, ϵ' are obtained from ϵ_{mono} and ϵ_{bis} by multiplying by $\log_e 10$. The values* of ϵ and ϵ' are shown in Table 1.

(f) *Extent of Direct Diazotization of Histidine*

The possibility of histidine reacting directly with nitrous acid to form nitroso- or diazo-compounds in the acid solution could not be neglected, particularly in view of the observations of Wiley and Lewis (1930) that casein treated

* A further series of detailed observations of extinction at various initial histidine concentrations has been made in an attempt to interpret the relationships in terms of the kinetic treatment put forward by Higgins and Williams (1952); the results suggest that the values for ϵ and ϵ' may require some modification.

with nitrous acid had lost some of its histidine. However, there is considerable evidence that the extent to which histidine takes part in such reactions is negligible in comparison with the coupling reaction and does not introduce spectrophotometric complications. The spectrum indicated by curve 4 of

TABLE 1

Compound	ϵ		ϵ'	
	380 m μ	490 m μ	380 m μ	490 m μ
Mono(azobenzenesulphonic acid)histidine	19,300*	3,300	44,400*	7,600
Bis(azobenzenesulphonic acid)histidine	16,200	22,800*	37,400	52,400*

* Maximum absorption.

Figure 4, which represents the standard procedure, using 1.02 mg. of histidine but omitting sulphanilic acid, absorbs very little in the visible range and the peak at *c.* 355 m μ , which still represents a small extinction compared with that shown by the coupled compounds, can be attributed to the ONO^- ion of nitrous acid. The extinction in the vicinity of 300 m μ , where the nitroso-group might be expected to absorb (Braude 1945), is very low. Again, since Beer's law is closely followed by the bisazo-compound (Fig. 6 (*a*)), when the corrections to the observed extinctions do not provide for the direct action of nitrous acid on the histidine, it appears unlikely that diazo- or nitroso-compounds are formed to any appreciable extent.

Further confirmation of this view was provided by the following experiment. The usual amounts of sulphanilic acid and sodium nitrite were mixed in the 25 ml. standard flask, histidine being omitted entirely in the acid stage. After 30 minutes sodium carbonate was added as before, followed by 8.52 mg. dissolved histidine and the solution was made up to volume with ethanol. The colour that developed denoted coupling, and the spectrum obtained is shown in Figure 7 against the spectrum given when a similar amount of histidine is present initially in the system and is thus subject to the action of nitrous acid. The spectra are similar in shape but less of the coupled compounds are formed when the histidine is added in the later (alkaline) stage. This is probably due to the conversion by the alkali of part of the diazonium compound to quaternary ammonium hydroxide and its subsequent conversion to an *isodiazotate* which does not participate in the coupling reaction. The relatively higher concentration of the monoazo-compound as indicated by the ratio of the extinction at 380 m μ to that at 490 m μ , follows logically from the higher ratio of histidine to diazonium salt (see above).

(g) Effect of Urea on Absorption Spectra

In order to eliminate directly the absorption due to unreacted nitrous acid within the system, some spectra were measured in the presence of urea, which destroys nitrous acid. One ml. sodium nitrite and 1 ml. 10 per cent. hydrochloric acid not containing sulphanilic acid were mixed in the 25 ml. flask and after 15 minutes 0.5 ml. 10 per cent. urea was added, which is slightly in excess of the equivalent required to destroy the nitrous acid. After another 15 minutes the system was made alkaline with sodium carbonate in the usual way, and after

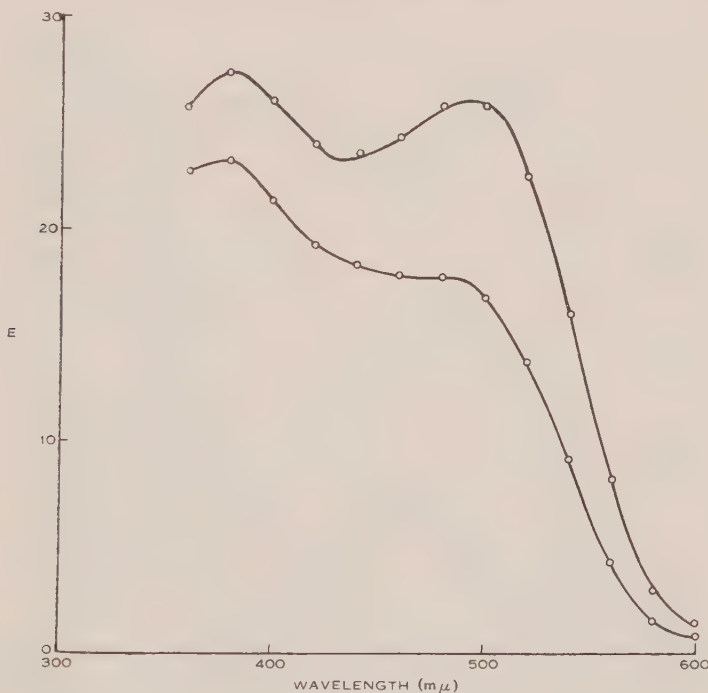


Fig. 7.—Absorption spectra when histidine is present initially in the system (upper curve) and when it is added in the alkaline stage (lower curve).

addition of ethanol made up to volume. The ONO^- peak at 355 mμ (Fig. 4, curve 2) was no longer present in the absorption spectrum, which is shown in Figure 8, curve 1. If sulphanilic acid and nitrous acid are present initially in the usual concentration and urea is again added after 15 minutes, followed by the alkali and ethanol after another 15 minutes, a bright yellow colour develops with maximum absorption at 430 mμ, as shown in curves 2 and 3 of Figure 8. When 1 ml. urea solution is used the extinction, and hence the concentration of the yellow compound, is lower than for 0.5 ml. urea. This appeared to be due to a deficiency of the diazonium salt. Histidine solution, containing 4.26 mg., was added to the final solution containing 1 ml. urea and the system was made up to twice the original volume. The resulting absorption spectrum, when corrected for dilution, was identical with that previously obtained (Fig. 8,

curve 2) showing that no coupling had taken place. The deficiency of diazonium salt suggests that diazotization of the sulphanilic acid was incomplete after the first 15 minutes. With the higher urea concentration more nitrous acid would be destroyed in the second 15 minutes, and less diazonium salt would be formed.

In the presence of histidine at high initial concentration coupling between histidine and the diazonium salt takes place preferentially as a practically instantaneous reaction and the formation of the yellow urea-diazonium compound

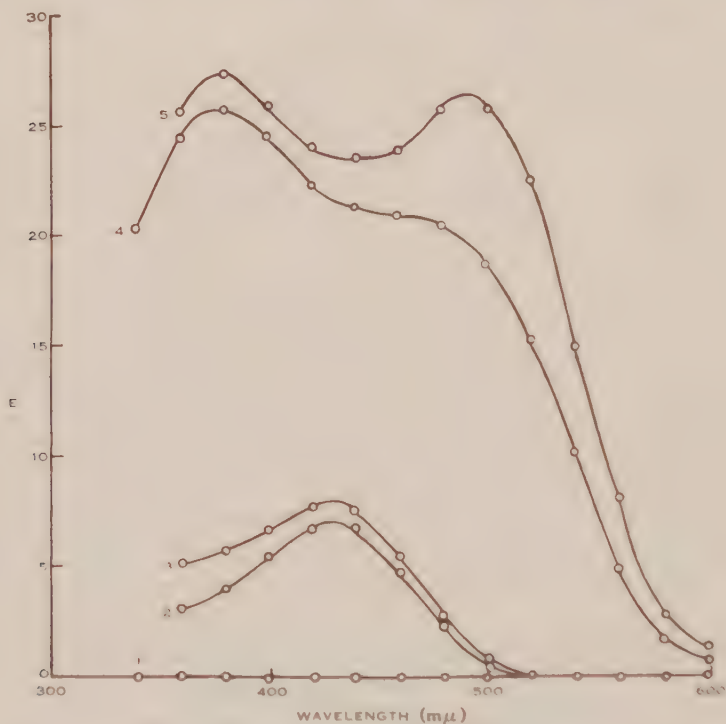


Fig. 8.—Effect of urea on absorption spectra. 1, Destruction of nitrous acid by 0.5 ml. urea (cf. Fig. 4); 2, absorption of urea-diazonium compound (1 ml. urea); 3, absorption of urea-diazonium compound (0.5 ml. urea); 4, absorption of histidine azo-compounds in presence of urea; and 5, absorption of histidine azo-compounds in absence of urea.

is not favoured. This is shown by the similarity between curves 5 and 4 of Figure 8, which represent respectively the standard procedure with 8.52 mg. histidine and the same with 1 ml. urea added 15 minutes before the sodium carbonate. The lower overall absorption in the presence of urea is again explicable in terms of a lower concentration of diazonium salt, and the higher absorption of the monoazo-compound at 380 mμ relative to that of the bisazo-compound at 490 mμ is consistent with the relatively higher histidine concentration.

(h) Absorption Spectra Observed with Tyrosine

The initial components of the system may be regarded as sulphanilic acid, nitrous acid, and tyrosine, and the reactions are closely analogous to those described for histidine, but with one important difference. The *o*-directing property of the phenolic hydroxyl group sponsors both direct reaction between the tyrosine and the nitrous acid and coupling between the tyrosine and the diazonium salt. The significance of the reaction between the tyrosine and the nitrous acid is that the occupation of an *o*-position in the ring would prevent

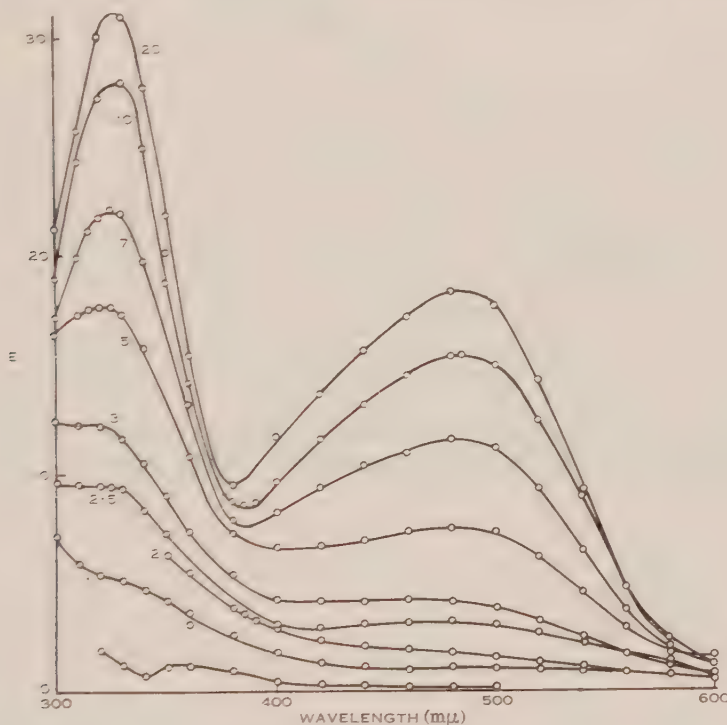


Fig. 9.—Absorption spectra of mono- and bis(azobenzenesulphonic acid)tyrosine at various initial tyrosine concentrations. Figures beside each curve are mg. tyrosine/25 ml. Lowest curve is spectrum obtained when system contains 10 mg. tyrosine but no sodium carbonate or ethanol.

subsequent bis-coupling. Absorption spectra observed for various tyrosine concentrations are shown in Figure 9. Mono(azobenzenesulphonic acid)tyrosine exhibits maximum absorption at 325 mμ and bis(azobenzenesulphonic acid)tyrosine at 480 mμ. Even at tyrosine concentrations well below the bis equivalent of diazonium salt, i.e. when $c_{\text{tyrosine}} < \frac{1}{2}c_{\text{sulph}}$, the monoazo-compound shows appreciable absorption. The indiscriminate formation of the mono- and bis-compounds is reflected in wide deviation from Beer's law in respect to uncorrected extinctions measured at the wavelengths of maximum absorption for the two compounds (Fig. 10).

(i) Preponderance of Mono Tyrosine Azo-Derivative

In order to determine whether the preponderance of the monoazo-compound, even at low initial tyrosine concentrations, was due to prior substitution in one

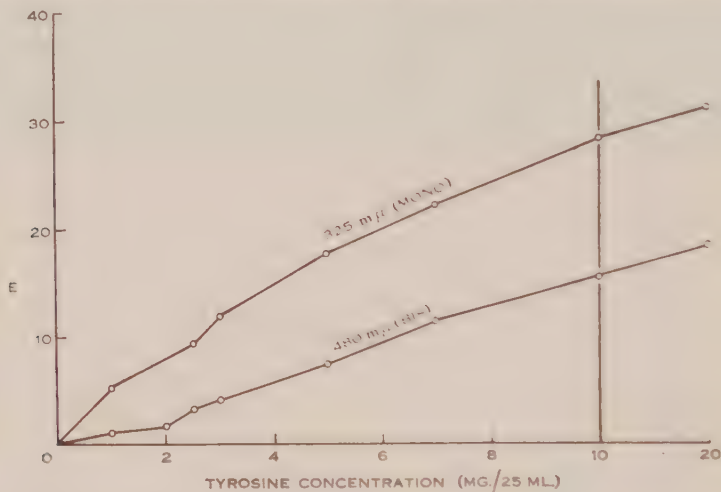


Fig. 10.—Dependence of final extinction E , at 325 and 480 $m\mu$, upon initial tyrosine concentration.

of the *o*-positions during the acid stage, the absorption spectrum was studied under conditions prohibiting direct diazotization of the tyrosine. To 1 ml. of

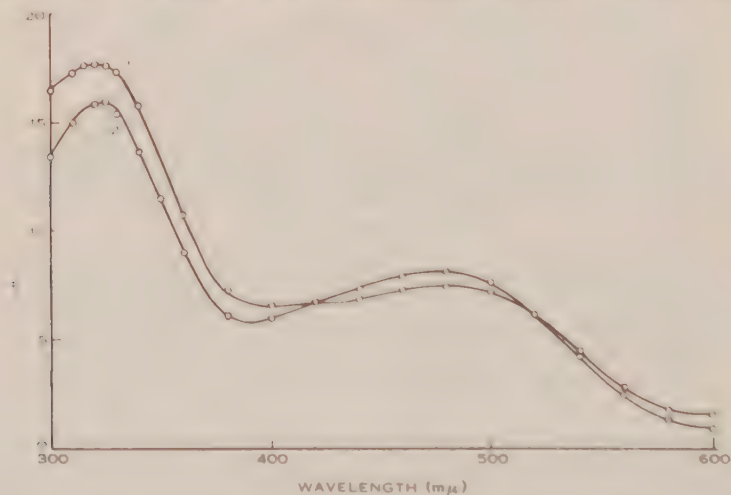


Fig. 11.—Effect of preventing direct diazotization of tyrosine on absorption spectra of coupled compounds. 1, Standard procedure (1 ml. sodium nitrite); and 2, sodium nitrite equivalent to sulphanilic acid (0.08 ml.) and tyrosine added later.

the sulphanilic acid was added only the equivalent amount of sodium nitrite, and, after 30 minutes, 5 mg. tyrosine was added and the solution made alkaline

as usual. The absorption spectrum obtained (Fig. 11) shows rather more bis-coupling and less mono-coupling than that obtained when direct diazotization is possible, indicating that *o*-diazotization does reduce the coupling potentiality to some extent. Other factors* must be involved, however, since both the mono- and the bisazo-derivatives are still formed at initial tyrosine concentrations below the equivalent of diazonium compound for bis-coupling.

The extinctions at 325 $m\mu$ and 480 $m\mu$ were observed in a standard system, containing 5 mg. tyrosine, at various times between $\frac{1}{2}$ hour and 20 hours after the addition of the alkali. Practically no change took place, showing that the observed effects were not due to lack of finality in the coupling reaction.

(j) *Direct Diazotization of Tyrosine*

The mechanism of direct diazotization has been studied by Philpot and Small (1938*a*) who postulate the initial formation of the nitroso-compound, followed by its rapid conversion by nitric oxide to the diazo-compound.

When tyrosine reacts with nitrous acid in the absence of sulphanilic acid, but with the usual amounts of sodium nitrite and hydrochloric acid, followed by sodium carbonate and ethanol, the absorption spectra are as shown in curves 1 and 2 of Figure 12 for 19 and 3 mg. tyrosine respectively. The absorption maximum at 355 $m\mu$ can again be ascribed to the $ONO-$ group. Herriott (1947) has observed an absorption peak at 390 $m\mu$ for diazotized tyrosine, but this is not apparent in curves 1 and 2 of Figure 12, possibly because diazotization was not sufficiently prolonged. The yellow colour characteristic of diazotized tyrosine deepens markedly over a period of days. From examination of the standard system in the absence of amino acids the wavelength of maximum absorption for the diazo-group was found to be well into the ultraviolet. The statement of Herriott (1947), that it absorbs in the vicinity of 400 $m\mu$, may arise from the work of Philpot and Small (1938*b*) on diazotized pepsin. These authors determined the extent of diazotization by measuring absorption at 411.5 $m\mu$, but did not suggest that this wavelength represented an absorption maximum.

The inflection between 400 and 450 $m\mu$ (curves 1 and 2 of Fig. 12) was absent when the system was not made alkaline, as shown by examination of a mixture containing 5 mg. tyrosine, sodium nitrite, and sulphanilic acid. This suggested that the inflection might be due to slight coupling between diazotized tyrosine and unreacted tyrosine itself.

(k) *Coupling between Tyrosine and Diazotized Tyrosine*

To examine further the possibility of diazotized tyrosine coupling with tyrosine itself, 10 mg. tyrosine was reacted with hydrochloric acid (not containing sulphanilic acid) and sodium nitrite for 30 minutes. The excess nitrous acid

* In the case of the imidazole ring of histidine, there are two virtually independent activated positions and since substitution in one will not affect the electron distribution at the other, bis substitution may be expected. However, in tyrosine there is only one activated unit and since *ortho*-substitution will tend to neutralize the electron disturbance due to the presence of the hydroxyl group, bis substitution will not be favoured.

was destroyed by 0.5 ml. urea solution and a further 10 mg. tyrosine was added. Upon the addition of sodium carbonate no colour developed, indicating that coupling was not pronounced, in accordance with the observations of Herriott (1947). However, an absorption peak appeared in the spectrum (Fig. 12, curve 3)

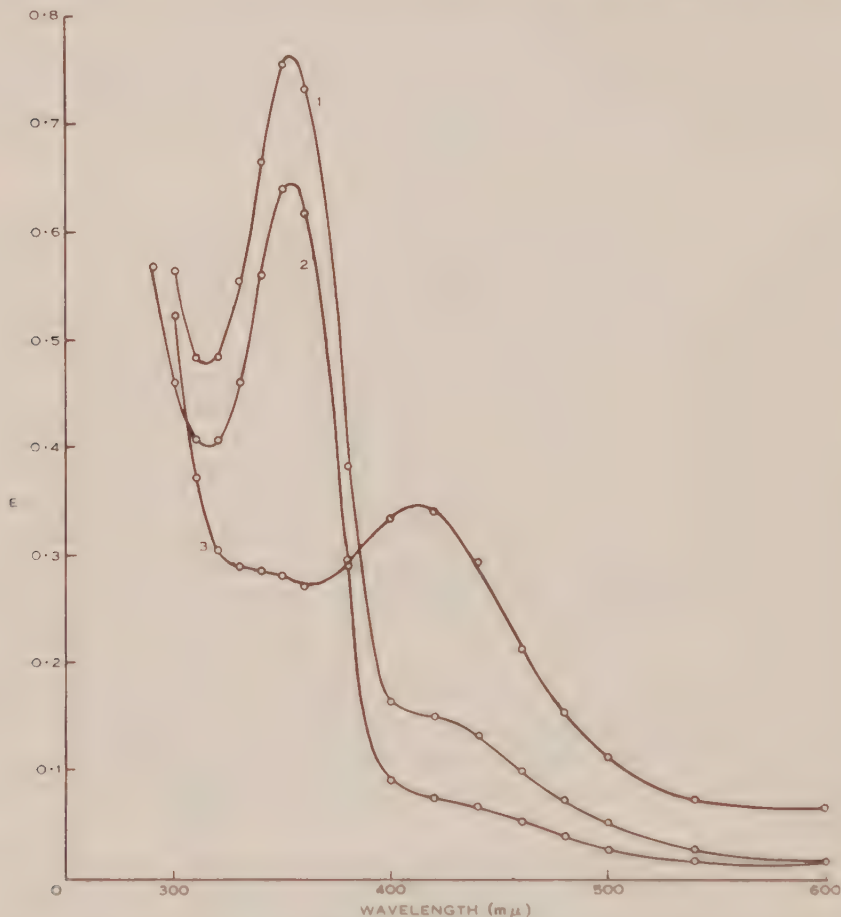


Fig. 12.—Action of nitrous acid on tyrosine. 1, 10 mg. tyrosine 25 ml. +hydrochloric acid (without sulphanilic acid) +sodium nitrite +sodium carbonate +ethanol; 2, as for curve 1, but with 3 mg. tyrosine 25 ml.; and 3, as for curve 1, +0.5 ml. urea after 15 minutes +further 10 mg. tyrosine just before addition of alkali.

at 410–415 mμ, in addition to the removal of the ONO^- band by the urea. It therefore seems probable that coupling takes place to a slight extent.

V. CONCLUSION

The investigation has revealed some of the conditions underlying the formation of mono- and bisazo-derivatives of histidine and tyrosine and the absorption characteristics of these compounds. Proper account of these results should be taken in applying the coupling reaction to the estimation of histidine

and tyrosine. In a later paper the question will be discussed further in relation to the spectra of various coupled proteins and protein hydrolysates.

VI. REFERENCES

- ALBANESE, A. A., DAVIS, V. I., SMETAK, E. M., LEIN, M., and FISHER, M. (1949).—*Arch. Biochem.* **20** : 47.
- ALBANESE, A. A., FRANKSTON, J. E., and IRBY, V. (1945).—*J. Biol. Chem.* **160** : 441.
- BOYD, W. C., and HOOKER, S. B. (1934).—*J. Biol. Chem.* **104** : 329.
- BOYD, W. C., and MOVER, P. (1935).—*J. Biol. Chem.* **110** : 457.
- BRAUDE, E. A. (1945).—*Rep. Progr. Chem.* **42** : 105.
- EAGLE, H., and VICKERS, P. (1936).—*J. Biol. Chem.* **114** : 193.
- HANKE, M. T. (1925).—*J. Biol. Chem.* **66** : 475.
- HANKE, M. T., and KOESSLER, K. K. (1920).—*J. Biol. Chem.* **43** : 527.
- HERRIOTT, R. M. (1947).—"Advances in Protein Chemistry." Vol. 3. p. 169. (Academic Press : New York.)
- HIGGINS, H. G., and WILLIAMS, E. J. (1952).—*Aust. J. Sci. Res. A* **5** : 572.
- JORGES, E. (1932).—*Biochem. J.* **26** : 1507.
- LANDSTEINER, K. (1945).—"The Specificity of Serological Reactions." (Harvard Univ. Press : Cambridge, Mass.)
- MACPHERSON, H. T. (1942).—*Biochem. J.* **36** : 59.
- PAULY, H. (1905).—*Hoppe-Seyl. Z.* **44** : 159.
- PAULY, H. (1915).—*Hoppe-Seyl. Z.* **94** : 284.
- PHILPOT, J. ST. L., and SMALL, P. A. (1938a).—*Biochem. J.* **32** : 534.
- PHILPOT, J. ST. L., and SMALL, P. A. (1938b).—*Biochem. J.* **32** : 542.
- PIPER, J. D., and BRODE, W. R. (1935).—*J. Amer. Chem. Soc.* **57** : 135.
- WILEY, F. H., and LEWIS, H. B. (1930).—*J. Biol. Chem.* **86** : 511.

NEW FLAVONES FROM *PONGAMIA PINNATA* (L.) MERR.

By L. RAMACHANDRA ROW*

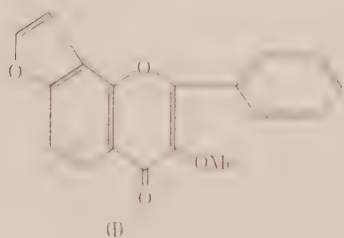
[Manuscript received July 24, 1952]

Summary

From the root bark of *Pongamia pinnata* (L.) Merr. (*P. glabra* Vent.) five flavones have been isolated. The two major constituents are shown to be karanjin and a new furanoflavone named *pongapin*. By degradation and synthesis the structure of the latter is established as 7,8-furano-3-methoxy-3',4'-methylenedioxyflavone. Of the three minor products, two appear to be isomeric with karanjin and pongapin respectively.

I. INTRODUCTION

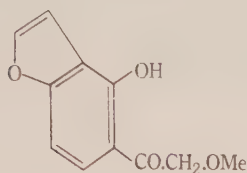
Varied physiological action has been reported for *Pongamia* spp. and the subject has been reviewed by Webb (1948). The plants have been used as fish poisons, and are also reported to be insecticidal and emetic. In India the seed oil and roots of *P. glabra* Vent. have long been reputed to cure certain skin diseases. Karanjin (I) was isolated from the oil by Limaye (1925, 1936, 1937) and the presence of a second unidentified compound has been reported by Rangaswami and Seshadri (1942). From the roots Rajagopalan *et al.* (1946) obtained kanugin (II). The structures of karanjin and kanugin were established through degradation and synthesis by Seshadri and Venkateswarlu (1941) and by Rao and Seshadri (1946). Until now karanjin has been the only known furanoflavone.



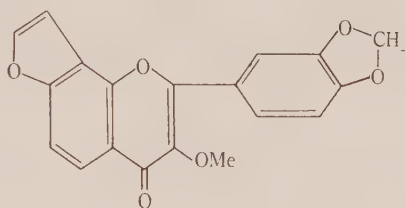
In the course of a survey of some Australian plants for possible insecticidal constituents, a detailed examination has been made of the root bark of *Pongamia pinnata* (L.) Merr. (*P. glabra* Vent.) grown in north Queensland. The dried bark was extracted with light petroleum and the products were separated by fractional crystallization, chromatography, and extraction with concentrated hydrochloric acid. Five crystalline products were isolated and these all give reactions characteristic of neutral flavones.

* Chemistry Department, University of Melbourne; present address: Chemistry Department, Andhra University, Waltair, India.

Compound *A* is identical with karanjin (I) since it gives no depression of melting point with an authentic specimen, gives the same colour reactions, and is degraded to the same ketone (III).



(III)



(IV)

Compound *B*, now named *pongapin*, has the composition $C_{19}H_{12}O_6$. It gives a positive reaction for the methylenedioxy-group, and on hydrolysis it yields the ketone (III) and piperonylic acid. It is therefore a methylenedioxykaranjin and its structure is confirmed as IV by synthesis from the ketone (III), piperonylic anhydride, and sodium piperonylate. Seshadri and Venkateswarlu (1941) established the structure of III by synthesis.

The ultraviolet absorption spectra of karanjin (I), kanugin (II), and pongapin (IV) have been measured in 95 per cent. ethanol using a Hilger "Uvispek" photoelectric spectrophotometer. The results are shown in Figure 1 and Table 1.

TABLE 1

Compound	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$\log \epsilon_{max}$
Karanjin ..	261	4.40			303	4.23
Kanugin ..	237*	4.32	320	4.25	341	4.32
Pongapin..	251.5	4.34			332.5	4.34

* Italicized figures refer to inflexions.

The absorption spectra show the marked effect of variations in the degree of substitution in the flavone nucleus. Karanjin, with no substituents in the phenyl group, shows considerable displacement of absorption bands in comparison with kanugin and pongapin. All three compounds show absorption characteristics which distinguish them from the quercetin and quercetagenin derivatives recently studied by Briggs and Locker (1951). These workers found three main bands but their band II is not found in the spectra of karanjin, kanugin, and pongapin, and the maximum corresponding to band III is found at lower wavelengths for these compounds. The furane ring does not appear to have any marked influence on the absorption. However, the inflexion found at 320 m μ in the kanugin spectrum, and at lower wavelengths and intensities in the quercetin and quercetagenin series, is not found in the corresponding regions of the spectra of the two furanoflavones.

The remaining three compounds have been obtained in small quantities only, and their examination will not be completed until more plant material is available.

Compound C, $C_{19}H_{12}O_6$, appears to be isomeric with pongapin, and like this compound it gives a test for the methylenedioxy-group, and yields piperonylic acid on hydrolysis. The phenolic fraction is a mixture and has not yet yielded any pure compound.

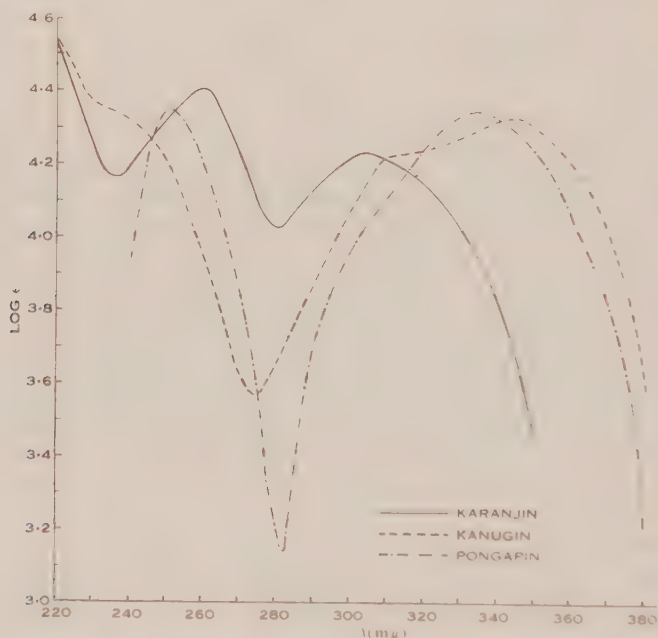


Fig. 1. Absorption spectra curves.

Compound D, $C_{18}H_{12}O_4$, appears to be isomeric with karanjin. The colour reaction with magnesium and hydrochloric acid is similar to that given by karanjin, suggesting an unsubstituted phenyl group. Confirmation of this is given by the isolation of benzoic acid by hydrolysis. The reaction does not proceed as smoothly as in the case of karanjin, and again the phenolic fraction is a mixture which is still under examination.

Compound E has not been investigated in detail. Its colour reactions and analysis suggest that it may be a more complex flavone than the other products.

II. EXPERIMENTAL

Melting points are corrected. Microanalyses by Dr. W. Zimmermann and assistants.

(a) Extraction and Isolation of the Products

Roots of *P. pinnata* were collected at Bingle Bay, Queensland. The bark was separated, and the milled material (750 g.) was continuously extracted with light petroleum (60 °C.). Removal of the solvent left a viscous yellow oil which did not crystallize. Ethanol (100 ml.) was added and the solution was allowed to stand in the refrigerator for 2 weeks. The precipitated solid

(0.25 g.) was collected, the ethanol distilled from the filtrate, and the residual oil mixed with benzene (50 ml.) and light petroleum (60 °C., 150 ml.). After standing for a further period of 2 weeks the precipitate (2.5 g.) was collected. The combined solids were fractionally crystallized from ethanol yielding compounds *A* and *B*. This mixture could also be separated by chromatography on alumina; benzene eluted compound *A*, and benzene-ethanol (1 : 1) gave compound *B*. The alcoholic mother liquors from which *A* and *B* separated were evaporated, the residue dissolved in benzene, and the solution passed over alumina. The column was then eluted with benzene-ethanol to give compound *C*.

The original benzene-light petroleum mother liquors were passed over alumina, and the column was developed with benzene, but evaporation of the eluate gave only oil. Elution with benzene-ethanol (1 : 1) then gave a solid, separated by crystallization from ethanol into compounds *C* and *D*. Since the remaining oils from the hydrocarbon solvents still gave tests for flavones they were dissolved in benzene and extracted several times with concentrated hydrochloric acid. The combined extracts were diluted with ice and water, and the pale yellow, gummy precipitate was fractionally crystallized from ethanol. The first fractions appeared to contain several compounds in small quantity, but only compound *D* (65 mg.) was obtained pure. The residue from the mother liquors was dissolved in benzene and passed over alumina, giving compound *E*.

(b) Examination of the Products

Compound *A* separated from ethanol in colourless rectangular plates, m.p. 157–158 °C. alone or mixed with authentic karanjin. Yield about 0.08%.

Found: C, 73.7; H, 4.0; OMe, 10.7%.

Calculated for $C_{18}H_{12}O_4$: C, 74.0; H, 4.1; OMe, 10.6%.

The product dissolved in concentrated sulphuric acid with a bright yellow colour, becoming blue-green on warming. Reduction in ethanol with magnesium and hydrochloric acid gave an orange colour. Hydrolysis with 8% ethanolic potassium hydroxide gave a phenolic ketone, m.p. 95–96 °C. alone or mixed with the compound similarly obtained from authentic karanjin.

Compound *B* (pongapin) formed very pale yellow needles from ethanol, m.p. 190–191 °C. Yield 0.08%.

Found: C, 67.7; H, 3.7; OMe, 9.6%.

Calculated for $C_{19}H_{12}O_6$: C, 67.9; H, 3.6; OMe, 9.2%.

With sulphuric acid pongapin gave a bright yellow colour, changing to blue (red in thick layers) on warming.

Sulphuric acid and gallic acid (test for methylenedioxy): green→blue.

Magnesium and hydrochloric acid: pink.

Compound *C* was sparingly soluble in ethanol, from which it separated in colourless silky needles, m.p. 232–233 °C. Yield 0.01%.

Found: C, 67.6; H, 3.8; OMe, 9.3%.

Calculated for $C_{19}H_{12}O_6$: C, 67.9; H, 3.6; OMe, 9.2%.

Sulphuric acid reaction: yellow→bright green.

Sulphuric acid and gallic acid: green→blue.

Magnesium and hydrochloric acid: pink.

Hydrolysis with 8% ethanolic potassium hydroxide gave piperonylic acid, but the phenolic fraction was a mixture containing coloured material.

Compound *D* was crystallized from ethanol (charcoal), forming colourless needles, m.p. 176–177 °C. Yield 0.02%.

Found: C, 74.2; H, 4.0; OMe, 10.8%.

Calculated for $C_{18}H_{12}O_4$: C, 74.0; H, 4.1; OMe, 10.6%.

Sulphuric acid reaction: yellow→green.

Sulphuric acid and gallic acid: green.

Magnesium and hydrochloric acid: orange-yellow.

Hydrolysis with 8% ethanolic potassium hydroxide gave benzoic acid, but the phenolic fraction contained colourless and coloured products not yet obtained pure.

Compound *E* was crystallized from ethanol as woolly needles, m.p. 134–136 °C. Yield 0.01%.

Found: C, 71.3; H, 4.3; OMe, 8.1%.

Sulphuric acid reaction: yellow→violet.

Sulphuric acid and gallic acid: green→blue.

Magnesium and hydrochloric acid: pink.

(c) Structure of Pongapin

(i) *Alkaline Hydrolysis*.—Pongapin (32 mg.) was heated under reflux for 8 hours with potassium hydroxide in absolute ethanol (8%, 2 ml.). The solvent was then removed under reduced pressure, and the residue was dissolved in water (15 ml.). Addition of hydrochloric acid gave a colourless precipitate. The suspension was extracted repeatedly with ether and the combined extracts were washed successively with 5% aqueous sodium bicarbonate (3 times) and 4% sodium hydroxide (4 times). The bicarbonate solution was acidified and continuously extracted with ether. Evaporation of the extract gave a white residue which was crystallized three times from water and then had m.p. 222.5–224 °C. (slight sintering at 218 °C.). No depression was obtained by addition of authentic piperonylic acid.

The sodium hydroxide extract was worked up in the same way, giving a pale yellow solid which, after crystallization from aqueous ethanol, formed colourless rectangular plates, m.p. 95–96 °C., not depressed by mixing with the ketone (III) obtained by similar hydrolysis of ketone (II).

Found: C, 63.9; H, 4.8; OMe, 15.6%.

Calculated for $C_{11}H_{10}O_4$: C, 64.0; H, 4.9; OMe, 15.1%.

(ii) *Synthesis*.—Piperonylic anhydride was first prepared by a modification of the method of Rao and Seshadri (1946). Piperonylic acid (3 g. dried at 110 °C. for 3 hours) was suspended in a mixture of dry ether (50 ml.) and pyridine (2 ml.) and cooled below 5 °C. Thionyl chloride (1.9 g.) was added in drops with stirring, and the mixture was kept at 0 °C. for 3 hours before pouring on to crushed ice. The white precipitate was filtered and washed successively with ice-cold dilute hydrochloric acid, ice-cold sodium bicarbonate, and ice-water. It was pressed between filter papers and finally dried over sulphuric acid in a vacuum. Yield 2.1 g. It crystallized from benzene in rectangular plates, m.p. 154–155 °C.

An intimate mixture of the ketone (III, 103 mg.), piperonylic anhydride (0.47 g.), and sodium piperonylate (0.2 g.) was heated for 3 hours at 170–180 °C. under reduced pressure. The solid was then refluxed with aqueous ethanolic sodium hydroxide (10%, 10 ml.) for 25 minutes. The solvent was removed under reduced pressure and the solid residue treated with water (40 ml.). The turbid, brown solution was extracted several times with ether, the extract was washed free of alkali and dried over magnesium sulphate. Evaporation left a solid which crystallized from ethanol in very pale yellow, shining needles, m.p. 190–191 °C., alone or mixed with natural pongapin. Yield 91 mg.

III. ACKNOWLEDGMENTS

The author is indebted to the administrators of the Commonwealth Technical Cooperation Scheme, and the Federal Government of Australia, for a fellowship which made this investigation possible, and to Professor E. J. Hartung for accommodation and facilities in the Chemistry Department, University of Melbourne. He also thanks Mr. R. G. Cooke, University of Melbourne, and Dr. J. R. Price, C.S.I.R.O., for encouragement, guidance, and helpful discussion. Mr. L. J. Webb, C.S.I.R.O., kindly supplied the plant material used in the work.

IV. REFERENCES

- BRIGGS, L. H., and LOCKER, R. H. (1951).—*J. Chem. Soc.* **1951**: 3141.
LIMAYE, D. B. (1925).—*Abstr. Indian Sci. Congr.* **1925**: 118, 126, 151.
LIMAYE, D. B. (1936).—*Rasayanam* **1**: 1.
LIMAYE, D. B. (1937).—*Rasayanam* **1**: 119.
RAJAGOPALAN, S., RANGASWAMI, S., RAO, K. V., and SESHADRI, T. R. (1946).—*Proc. Indian Acad. Sci.* **23A**: 60.
RANGASWAMI, S., and SESHADRI, T. R. (1942).—*Proc. Indian Acad. Sci.* **15A**: 417.
RAO, K. V., and SESHADRI, T. R. (1946).—*Proc. Indian Acad. Sci.* **23A**: 147.
SESHADRI, T. R., and VENKATESWARLU, V. (1941).—*Proc. Indian Acad. Sci.* **13A**: 404.
WEBB, L. J. (1948).—*Coun. Sci. Indust. Res. Aust. Bull. No. 232*, p. 89.

COLOURING MATTERS OF AUSTRALIAN PLANTS

II. NAPHTHOQUINONES FROM *DIOSPYROS HEBECARPA* A. CUNN.

By R. G. COOKE* and H. DOWD*

[Manuscript received May 16, 1952]

Summary

Ether extraction of *Diospyros hebecarpa* A. Cunn. yields plumbagin, and two new compounds which are shown to be 5-hydroxy-7-methyl-1,4-naphthoquinone and 1,4-diketo-5-hydroxy-7-methyl-1,2,3,4-tetrahydronaphthalene. The ultraviolet absorption spectra of these and related compounds are recorded, and the synthesis of 1,4,5-trimethoxy-7-methylnaphthalene is described.

I. INTRODUCTION

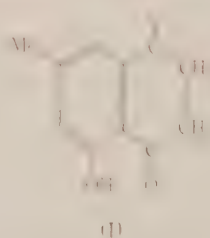
About a dozen species of *Diospyros* L. have been recorded in eastern Australia. *D. hebecarpa* A. Cunn. is a small or medium tree common on parts of the tropical coast of Queensland and extending northwards into New Guinea and the neighbouring islands. Webb (1948) records that the fresh fruits have been used as a fish poison by the aborigines and that the plant has also been suspected of poisoning livestock.

Toxic and vesicant properties have been attributed to related species in other countries, and plumbagin (a known vesicant) has been isolated by Mejer (1947) from *D. maritima*, and by Paris and Moysé Mignon (1949) from *D. guineensis* and *D. mespiliformis*.

A brief report of the isolation of the constituents of *D. hebecarpa* has been given by Cooke, Dowd, and Webb (1952) and the chemical investigation of these compounds is now described in detail.

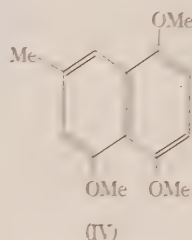
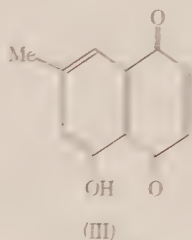
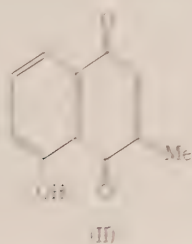
Cold ether extraction of the bark, and of some samples of leaves and fruits, has yielded plumbagin, 2-methyl-5-hydroxy-1,4-naphthoquinone, which was identified by comparison with an authentic specimen isolated from *Plumbago zeylanica*.

From the fresh leaves, and from some fresh immature fruits, a new colourless compound $C_{11}H_{10}O_3$ has been obtained. The following evidence indicates that this is 1,4-diketo-5-hydroxy-7-methyl-1,2,3,4-tetrahydronaphthalene (I).



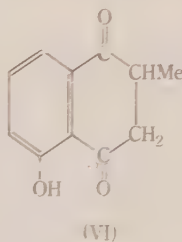
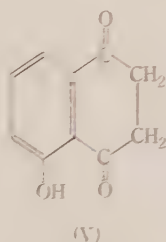
* Chemistry Department, University of Melbourne.

Acetylation gives a *triacetate* and methylation with dimethyl sulphate gives a *trimethyl ether*. Oxidation yields a *quinone* which is isomeric with plumbagin and shows the colour reactions and absorption spectrum characteristic of naphthoquinones of the juglone type. This quinone is not identical with the other known methyljuglone (II) which was prepared by de Buruaga (1933), and by Fieser and Dunn (1936). Its structure is established as 7-methyljuglone (III) by synthesis of 1,4,5-trimethoxy-7-methylnaphthalene (IV), which is identical with the trimethyl ether obtained from I.



The quinone (III) has been isolated from one sample of leaves, which contained no plumbagin, and is also formed by exposing the sliced fresh fruit to the air. It is noteworthy that the structural features of this naphthoquinone are found in a number of naturally occurring anthraquinones, e.g. chrysophanic acid, emodin, and helminthosporin.

The diketone-structure of I is supported by the close resemblance of this compound to 3-hydrojuglone (V) and 3-hydroplumbagin (VI), the structures of which were established recently by Thomson (1950, 1951).



Like these compounds, I is volatile in steam and is not affected by diazo-methane. It also shows the same characteristic ultraviolet absorption.

The isolation of I apparently gives the first well-authenticated example of the natural occurrence of a diketotetrahydronaphthalene. Mylius (1885) claimed to have isolated 3-hydrojuglone (V) from walnuts but Daglish (1950) was unable to confirm this and has suggested that the original method of isolation could have converted α -hydrojuglone to the keto-isomer. No such conversion seems likely in the isolation of I.

The absorption spectra of all these compounds were measured in 95 per cent. ethanol with a Hilger "Uvispek" photoelectric spectrophotometer. The results are given in Table 1 and Figures 1 and 2.

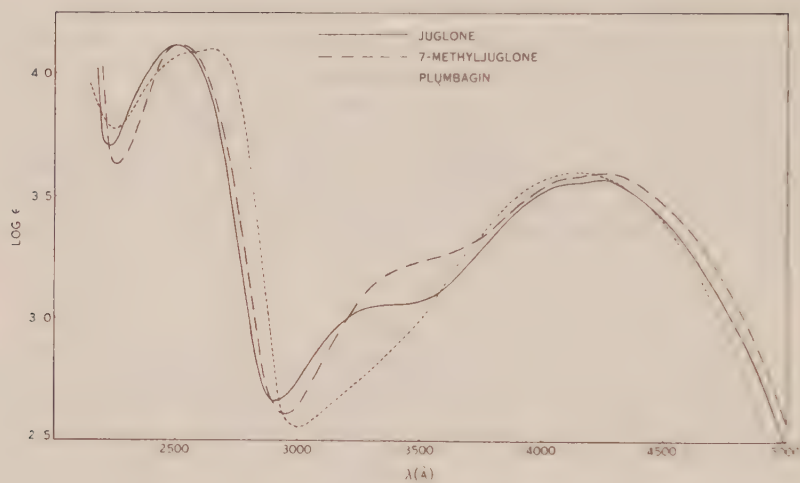


Fig. 1

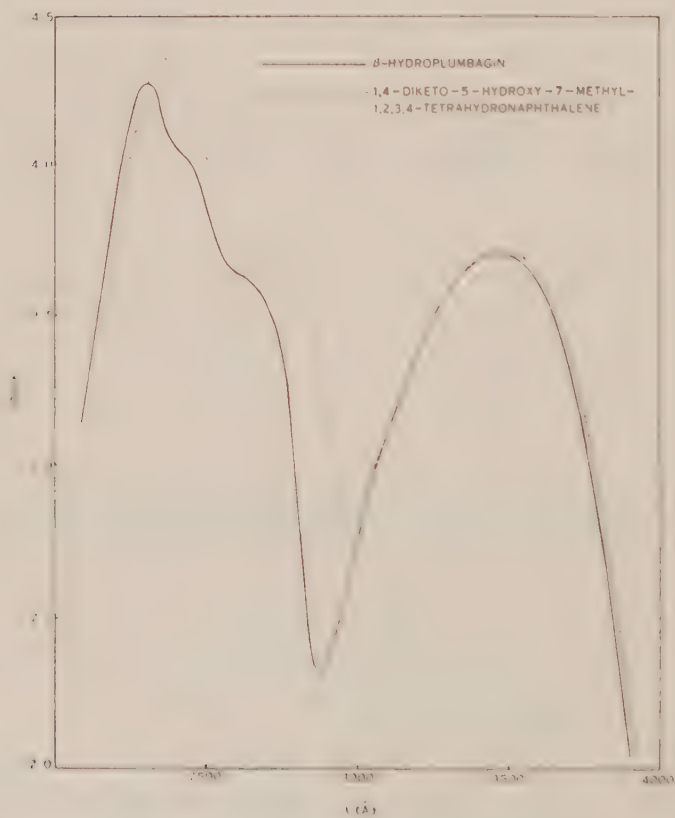


Fig. 2

TABLE I

Compound	λ_{max} (Å)	$\log \epsilon_{max}$	λ_{max} (Å)	$\log \epsilon_{max}$	λ_{max} (Å)	$\log \epsilon_{max}$	λ_{max} (Å)	$\log \epsilon_{max}$
1, 4 - Diketo - 5 - hydroxy-7-methyl- 1,2,3,4-tetrahydro- naphthalene (I)	2355	4.36	2475*	4.05	2700	3.77	3475	3.73
β -Hydrojuglone ..	2300	4.33	2430	4.08	2600	3.73	3465	3.71
β -Hydroplumbagin	2300	4.28	2430	4.04	2600	3.63	3475	3.71
Juglone			2490	4.12	3400	3.05	4250	3.57
Plumbagin ..			2650	4.10			4150	3.60
3-Methyljuglone ..			2650	4.07			4175	3.59
7-Methyljuglone ..			2525	4.12	3500	3.26	4250	3.60

* Italicized figures refer to inflexions.

II. EXPERIMENTAL

Unless otherwise stated, melting points are corrected and were observed in Pyrex capillaries. Light petroleum refers to the fraction of b.p. 40–60 °C.

(a) Extraction of Plant Materials

(i) *Fruits*.—The fresh fruits were sliced and allowed to soak in ether for several days. The ether solution was then separated and evaporated, the residue being dried in a vacuum desiccator. The dried product was extracted with boiling light petroleum and the filtered solution was concentrated to give the crystalline product. A sample of immature fruits collected from a tree on the mountain slopes near Cairns in September 1949 gave only 1,4-diketo-5-hydroxy-7-methyl-1,2,3,4-tetrahydronaphthalene (I). More mature fruits collected from Rocky Point, near Cairns, in November 1951 gave only plumbagin.

(ii) *Leaves*.—The fresh leaves were extracted in the same way, but the residue from the ether solution was purified by steam distillation, fractional crystallization from light petroleum (in which the plumbagin is much more soluble), and sublimation. One sample of leaves collected from a small, stunted tree in open country near Cairns in April 1949 gave only 7-methyljuglone. Another sample obtained from a large tree in the same area during September 1949 gave plumbagin and 1,4-diketo-5-hydroxy-7-methyl-1,2,3,4-tetrahydronaphthalene (I).

(iii) *Bark*.—The coarsely crumbled bark was also extracted by steeping in ether. The residue on evaporation was then treated with warm ethanol which dissolved the plumbagin. The remaining sparingly soluble material could be crystallized from acetone in colourless plates, m.p. above 260 °C. It has not been investigated further.

(b) Properties of Products

(i) *Plumbagin*.—This compound was sublimed under reduced pressure and crystallized repeatedly from aqueous ethanol to form long, fine, orange needles, m.p. 76–77 °C. Fieser and Dunn (1936) report m.p. 78–79 °C.; Thomson (1951) gives m.p. 77 °C.

Found: C, 70.2; H, 4.4%.

Calculated for $C_{11}H_8O_3$: C, 70.2; H, 4.3%.

The leucotriacetate was prepared by reductive acetylation in the usual way. It formed colourless prisms from chloroform-light petroleum, m.p. 124–125 °C.

An authentic specimen of plumbagin was isolated by extraction of the ground root of *P. zeylanica* with light petroleum. After sublimation and crystallization from aqueous ethanol it gave fine orange-yellow needles, m.p. 76–77 °C. alone or mixed with the above product from

D. hebecarpa. The leucoacetate also gave the same melting point and mixed melting point. Plumbagin regenerated from the purified leucoacetate by hydrolysis and oxidation showed no change in properties.

(ii) *1,4-Diketo-5-hydroxy-7-methyl-1,2,3,4-tetrahydronaphthalene* (I). — Repeated crystallization of this compound from light petroleum gave very pale yellow plates or needles, m.p. 112–113 °C. Further crystallization from aqueous ethanol containing a trace of sodium dithionite gave small colourless needles having the same melting point. The compound is volatile in steam.

Found: C, 69.6; H, 5.3%.

Calculated for $C_{11}H_{10}O_3$: C, 69.5; H, 5.3%.

The triacetate was prepared by heating for 1 hour with acetic anhydride and sodium acetate. It crystallized from ethanol as colourless plates, m.p. 175.5–176.5 °C.

Found: C, 64.4; H, 5.1%.

Calculated for $C_{17}H_{16}O_6$: C, 64.5; H, 5.1%.

The trimethyl ether was obtained by heating with dimethyl sulphate and sodium hydroxide in the presence of a little sodium dithionite. It was purified by sublimation in a vacuum and crystallization from light petroleum. It was obtained as small plates or needles, m.p. 112–113 °C.

Found: C, 72.1; H, 7.1%.

Calculated for $C_{14}H_{16}O_3$: C, 72.4; H, 7.0%.

Only unchanged material was recovered from attempted methylation with diazomethane.

(iii) *7-Methyljuglone* (III).—This compound was isolated from the leaves or prepared from I by adding ferric chloride to an aqueous ethanol solution containing hydrochloric acid. It was separated by distillation in steam and, after drying in ether, it was further purified by sublimation *in vacuo* and crystallization from light petroleum. It formed elongated red prisms, m.p. 125.5–126.5 °C. (In soda glass the compound melted with extensive decomposition in the range 90–110 °C.)

Found: C, 69.9; H, 4.3%.

Calculated for $C_{11}H_8O_3$: C, 70.2; H, 4.3%.

This quinone gives an unstable violet colour with aqueous alkali, a clear mauve solution with dilute aqueous nickel acetate, a brilliant red colour with concentrated sulphuric acid, changing to dull purple on heating, and a red colour with Dimroth's reagent, changing to purple-red on warming.

The acetate was prepared by adding a trace of concentrated sulphuric acid to a solution of the quinone in the minimum quantity of acetic anhydride at 0° C. After standing for 20 minutes water was added and the yellow product was crystallized from chloroform-light petroleum giving fine yellow needles, m.p. 151–152 °C.

Found: C, 67.6; H, 4.4%.

Calculated for $C_{13}H_{10}O_4$: C, 67.8; H, 4.4%.

The leucotriacetate, obtained by reductive acetylation, crystallized from ethanol in colourless plates, m.p. 175.5–176.5 °C. alone or mixed with the product obtained by direct acetylation of I.

(c) *Synthesis of 1,4,5-Trimethoxy-7-methylnaphthalene*

(i) *2,5-Dimethoxypropionophenone*.—Anhydrous aluminium chloride (72 g.) was added over a period of 1½ hours to a stirred mixture of 1,4-dimethoxybenzene (58 g.) and propionyl chloride (48 g.) in tetrachloroethane (80 ml.) and nitrobenzene (20 ml.). The temperature was kept below 0 °C. and the mixture was protected from atmospheric moisture. After standing 6 hours in a freezing mixture the apparatus was transferred to a refrigerator for 4 days. After decomposition with ice and hydrochloric acid and distillation of the solvents in steam the residual oil was dissolved in ether and washed with sodium hydroxide (10%). The recovered oil was distilled giving a faintly yellow liquid, b.p. 102–103 °C./0.3 mm. Yield 85%. The product gave no

colour with ferric chloride solution. This preparation seems to be superior to that described by Johnson and Hodge (1913) (see Cruickshank and Robinson 1938).

The semicarbazone crystallized from ethanol in fine needles, m.p. 191–192 °C.

Found: N, 16.8%.

Calculated for $C_{12}H_{17}O_3N_3$: N, 16.7%.

The 2,4-dinitrophenylhydrazone crystallized from acetone-ethanol in fine, bright red needles, m.p. 173–174 °C.

Found: N, 14.9%.

Calculated for $C_{17}H_{18}O_6N_4$: N, 15.0%.

Two by-products were recovered from the alkaline washings of this ketone. One of these was found to be 2-hydroxy-5-methoxypropio-phenone, m.p. 47–48 °C. Auwers and Müller (1917) give m.p. 47–49 °C. The other is apparently formed by the introduction of a second propionyl group. It formed fine yellow needles from ethanol, m.p. 123.5–124 °C.

Found: C, 66.2; H, 6.7; OMe, 13.2%.

Calculated for $C_{13}H_{16}O_4$: C, 66.1; H, 6.8; OMe, 13.1%.

(ii) *α -Bromo-2,5-dimethoxypropio-phenone*.—Bromine (40 g.) in carbon tetrachloride (50 ml.) was added slowly to a solution of 2,5-dimethoxypropio-phenone (50 g.) in carbon tetrachloride (100 ml.). Removal of the solvent under reduced pressure left a reddish oil which crystallized on standing. It was recrystallized from ethanol in yellow prisms, m.p. 47–48 °C.

Found: C, 48.3; H, 4.8; OMe, 22.7%.

Calculated for $C_{11}H_{13}O_3Br$: C, 48.4; H, 4.8; 2×OMe, 22.7%.

(iii) *Ethyl α -Carbethoxy- β -(2,5-dimethoxybenzoyl)butyrate*.—The above bromoketone, dissolved in absolute ethanol, was added slowly at 0 °C. to 1 mole of sodio-malonic ester in ethanol. After standing 24 hours the mixture was boiled under reflux for 30 minutes. Water was added, the product was dried in ether, and after removing unreacted malonic ester by distillation under reduced pressure it crystallized from ethanol in colourless prisms, m.p. 57–58 °C. Yield 44% in the two steps from 2,5-dimethoxypropio-phenone.

Found: C, 61.5; H, 6.7%.

Calculated for $C_{18}H_{21}O_7$: C, 61.4; H, 6.9%.

(iv) *β -Methyl- γ -(2,5-dimethoxyphenyl)butyric Acid*.—The malonic ester derivative (28 g.) in ethanol (240 ml.) was added to freshly amalgamated zinc (150 g.) and concentrated hydrochloric acid (200 ml.). The mixture was boiled under reflux for 8 hours, hydrochloric acid (10 ml.) being added after every 2 hours. The product was isolated by distillation to half-volume, addition of water, and extraction with ether. The light orange oil was hydrolysed by refluxing for 7 hours with aqueous-ethanolic potash containing some sodium dithionite. After 2 hours' heating dimethyl sulphate (5 ml.) was slowly added through the condenser. The mixture was cooled, acidified with concentrated hydrochloric acid, and extracted with ether. Distillation left a brown oil which was heated to 150 °C. to cause decarboxylation. When the effervescence subsided the temperature was raised to 180 °C. for a short time. An ethereal solution of the viscous oil was extracted with aqueous sodium carbonate, and the brown oil liberated on acidification was recovered with ether and purified by distillation under reduced pressure. It formed a viscous greenish oil, b.p. 190 °C./0.5 mm.

(v) *3-Methyl-5,8-dimethoxytetralone-1*.—The above acid (3 g.) was heated for 10 minutes at 60 °C. with a mixture of water (6 ml.) and concentrated sulphuric acid (24 ml.). The mixture was poured into ice-water and the product extracted with ether. After washing with sodium bicarbonate, drying, and evaporating, the residue was purified by sublimation in a vacuum and crystallization from light petroleum. It formed colourless needles, m.p. 74.5–75.5 °C.

Found: C, 70.7; H, 7.1%.

Calculated for $C_{13}H_{16}O_3$: C, 70.9; H, 7.3%.

(vi) *3-Methyl-5,8-dimethoxy-1-naphthol*.—The tetralone (0.29 g.) and 20% palladium charcoal (0.03 g.) were refluxed in *p*-cymene (5 ml.) for 10 hours. Ether was added, the catalyst was removed by filtration, and the solution was extracted with 10% sodium hydroxide. No product was extracted because the naphthol is a cryptophenol. However, removal of the ether and cymene left a crystalline residue. This was purified by sublimation under reduced pressure and crystallization from ethanol. It formed fine needles, m.p. 117–118 °C. The substance was not soluble in aqueous alkali, but it coupled with diazonium solutions in aqueous-ethanolic alkali. Attempts to demethylate this compound gave no definite product.

Found: C, 71.3; H, 6.4; OMe, 27.9%.

Calculated for $C_{13}H_{14}O_3$: C, 71.5; H, 6.5; 2 × OMe, 28.4%.

(vii) *1,4,5-Trimethoxy-7-methylnaphthalene*.—The above naphthol was methylated with dimethyl sulphate and sodium hydroxide in aqueous acetone. The product was purified by sublimation in a vacuum and crystallization from light petroleum. It formed small plates or needles, m.p. 112–113 °C. alone or mixed with the trimethyl ether obtained from the natural product I.

(d) Preparation of other Compounds for Spectrophotometry

(i) *Juglone*.—This was prepared according to the directions of Willstätter and Wheeler (1914) and was purified by sublimation in a vacuum and crystallization from light petroleum (b.p. 60–70 °C.). It formed bright orange needles, m.p. 164–165 °C. (In soda glass the compound melted with extensive decomposition in the range recorded by many previous workers.)

(ii) *3-Methyljuglone*.—A modification of the method of Fieser and Dunn (1936) was used for the preparation of this quinone: hydrochloric acid was added to hydrolyse the acetate during steam distillation, and, after vacuum sublimation and crystallization from ethanol, the product formed light orange needles, m.p. 160–161 °C. Fieser and Dunn (1936) give m.p. 157–158 °C.

(iii) *β-Hydrojuglone*. This substance was obtained by the method of Willstätter and Wheeler (1914) and was purified by sublimation, crystallization from aqueous ethanol containing a trace of sodium dithionite, and finally from light petroleum. It formed colourless plates, m.p. 94–95 °C. Mylius (1885) reports colourless plates or needles, m.p. 96–97 °C., and Thomson (1950) gives pale yellow needles or plates, m.p. 96–97 °C.

(iv) *β-Hydroplumbagin*.—The method of Madinaveitia and Olay (1933) was used and the product was purified as for β-hydrojuglone. It formed fine colourless needles, m.p. 87–88 °C. Madinaveitia and Olay (1933) obtained yellow prisms, m.p. 86 °C., and Thomson (1951) reported pale yellow needles, m.p. 87 °C.

III. ACKNOWLEDGMENTS

We are indebted to Mr. L. J. Webb, C.S.I.R.O. Plant and Soils Laboratory, Brisbane, for arranging the collection of the plant material, and for information about its occurrence and properties. We also thank the Dafydd Lewis Trust for a scholarship held by H. Dowd during the course of this work. Microanalyses were done by Dr. W. Zimmermann and staff.

IV. REFERENCES

- AUWERS, K. v., and MÜLLER, W. (1917).—*Ber. dtsch. chem. Ges.* **50**: 1175.
 COOKE, R. G., DOWD, H., and WEBB, L. J. (1952).—*Nature* **169**: 974.
 CRICKSHANK, J. H., and ROBINSON, R. (1938).—*J. Chem. Soc.* **1938**: 2064.
 DAGLISH, C. (1950).—*Biochem. J.* **47**: 452.
 DE BURUAGA, J. S. (1933).—*An. Soc. Esp. Fis. Quim.* **31**: 185.
 FIESER, L. F., and DUNN, J. T. (1936).—*J. Amer. Chem. Soc.* **58**: 572.
 JOHNSON, T. B., and HODGE, W. W. (1913).—*J. Amer. Chem. Soc.* **35**: 1014.
 MADINAVEITIA, A., and OLAY, E. (1933).—*An. Soc. Esp. Fis. Quim.* **31**: 134.

- MEIJER, Th. M. (1947).—*Rec. Trav. Chim. Pays-Bas* **66** : 193.
- MYLIUS, F. (1885).—*Ber. dtsh. chem. Ges.* **18** : 2567.
- PARIS, R., and MOYSE-MIGNON, H. (1949).—*C.R. Acad. Sci. Paris* **228** : 2063.
- THOMSON, R. H. (1950).—*J. Chem. Soc.* **1950** : 1737.
- THOMSON, R. H. (1951).—*J. Chem. Soc.* **1951** : 1237.
- WEBB, L. J. (1948).—*Coun. Sci. Industr. Res. Aust. Bull.* No. 232, p. 50.
- WILLSTÄTTER, R., and WHEELER, A. S. (1914).—*Ber. dtsh. chem. Ges.* **47** : 2796.

ALKALOIDS OF THE AUSTRALIAN RUTACEAE: *PENTACERAS AUSTRALIS* HOOK. F.

III. IDENTIFICATION OF 4-METHYLTHIOCANTHIN-6-ONE

By EVA R. NELSON* and J. R. PRICE†

[Manuscript received May 6, 1952]

Summary

The alkaloid $C_{15}H_{10}OSN_2$ occurring in the bark of *Pentaceras australis* Hook. f. is shown to be 4-methylthiocanthin-6-one. The lactam ring is opened by alcoholic alkali giving 2-methylthio-2-(1'- β -carboly)acrylic acid, followed by the elimination of methyl mercaptan and recyclization to 4-hydroxycanthin-6-one. Confirmation of the relationship of the alkaloid to canthinone is provided by (i) the ultraviolet absorption spectra, (ii) the formation of methyl mercaptan and 4,5-dihydrocanthine by treatment with zinc and hydrochloric acid, and (iii) the formation, by treatment with Raney nickel, of canthinone, 4,5-dihydrocanthinone, or the methyl or ethyl esters of β -carbolypropionic acid, depending on the conditions employed. The alkaloid has been synthesized from β -carboline-1-carboxylic acid by conversion of the acid chloride with malonic ester, followed by cyclization, to 4-hydroxycanthinone, thence to 4-chlorocanthinone which with potassium methyl mercaptide yields 4-methylthiocanthinone. The low basicity of the alkaloid is discussed, particularly in relation to the basicity of 4-methoxycanthinone.

I. THE STRUCTURE OF THE ALKALOID

The bark of *Pentaceras australis* Hook. f. contains three alkaloids, two of which have been identified as canthin-6-one and 5-methoxycanthin-6-one (Haynes, Nelson, and Price 1952; Nelson and Price 1952). The third alkaloid $C_{15}H_{10}OSN_2$ is now shown to be 4-methylthiocanthin-6-one (I). It is attacked slowly by ethanolic alkali and more rapidly by amyl alcoholic alkali with the formation of a yellow acidic compound $C_{14}H_8O_2N_2$, which by oxidation with permanganate gives the known β -carboline-1-carboxylic acid. The acidic substance forms an acetyl derivative $C_{14}H_7ON_2OOC(CH_3)$, transformable with moist ethereal diazomethane to a monomethoxy-derivative; the ultraviolet absorption spectra of both the acetyl derivative and the methoxy-compound (see Fig. 1 and Table I) closely resembling those of canthin-6-one, 5-acetoxycanthin-6-one and 5-methoxycanthin-6-one. Unlike 5-hydroxycanthinone the substance $C_{14}H_8O_2N_2$ does not react with α -phenylenediamine, is not reduced by Raney alloy sodium hydroxide, and gives a red colour with ferric chloride. It is, therefore, 4-hydroxycanthin-6-one (II), a structure compatible with the Fe^{+++} reaction and with the greater resistance to hydrogenation by Raney

* Chemistry Department, University of Melbourne; present address: Nicholas Pty. Ltd. Melbourne.

† Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

alloy-sodium hydroxide as compared with 5-hydroxycanthinone. This identification has been confirmed by synthesis. β -Carboline-1-carboxylic acid chloride, which, contrary to the statement of Snyder, Walker, and Werber (1949), we found to be readily prepared from the acid, was condensed with the magnesium ethoxy-derivative of malonic ester giving a product which after acid hydrolysis yielded 4-hydroxycanthinone.

TABLE I
ULTRAVIOLET ABSORPTION DATA

Substance	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$\log \epsilon_{max}$	λ_{max} (m μ)	$\log \epsilon_{max}$
4-Methoxycanthin-6-one	369 352 ~338	4.20 4.15 3.87	293 284	4.09 4.10	266 ~238 ~226	4.32 4.34 4.43
4-Acetoxycanthin-6-one	380 362 ~346 ~333	4.10 4.11 3.88 3.68	299 289	3.92 3.93	267 259	4.18 4.15
Canthin-6-one*	381 362 347	4.14 4.17 3.94	299 ~293	3.91 3.90	269 259 251	4.03 4.05 4.09
4-Methylthiocanthin-6-one	380 362 ~350	4.14 4.19 4.03	306 294	4.09 4.22	253 239	4.41 4.42
Methyl ester of 2-methylthio-2-(1'- β -carboly)acrylic acid	351 ~340	3.81 3.75	290	4.35	235	4.49
Methyl ester of 2-(1'- β -carboly)acrylic acid*	382	3.94	297	4.30	~240	4.33
β -Carboline†	342 ~328	3.8 3.7	285	4.4	220-230	4.7

* Haynes, Nelson, and Price (1952).

† Horner (1939).

The second product of the action of alcoholic alkali on the alkaloid is methyl mercaptan, identified by means of its 2,4-dinitrophenyl derivative and its mercury salt in the gases from the amyl alcoholic reaction mixture.* When only a short time is allowed for the reaction between amyl alcoholic potash and

* Despite the mercaptan-like odour of the gases evolved during the reaction with ethanolic potash, the presence of methyl mercaptan could not be demonstrated, presumably because the reaction rate is so slow that side reactions such as oxidation destroy the greater part of the mercaptan. Increasing the reaction rate by the use of amyl alcohol overcomes the difficulty.

the alkaloid, the lactam ring is opened without hydrolysis of the thioether group and the product is a colourless acid $C_{15}H_{12}O_2SN_2$ which gives a methyl

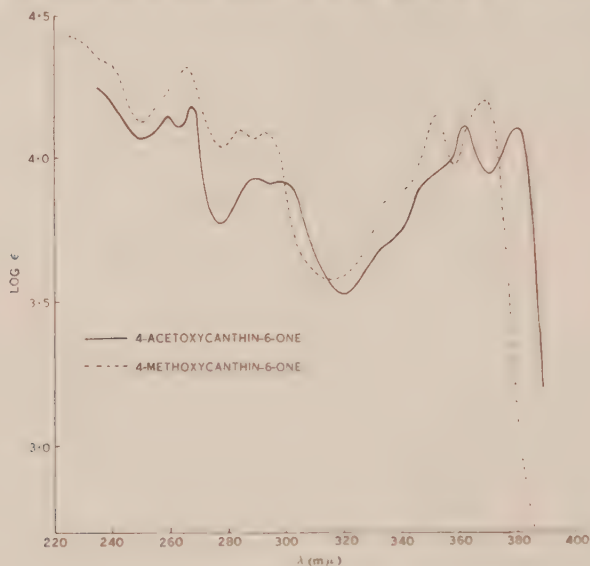


Fig. 1

ester $C_{16}H_{14}O_2SN_2$. The ultraviolet absorption spectrum of the methyl ester (see Fig. 2 and Table 1) closely resembles that of β -carboline. The acid is

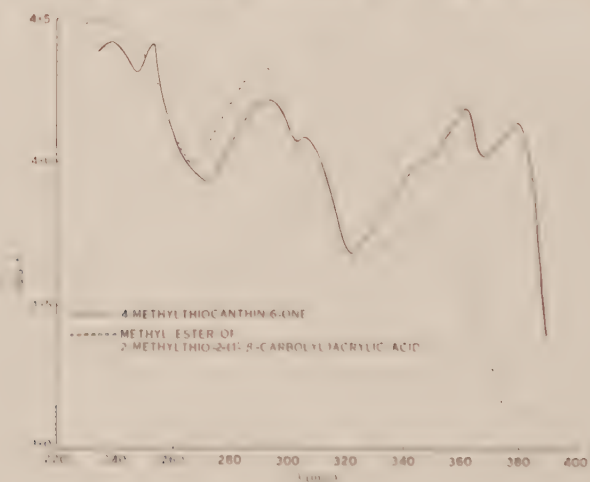
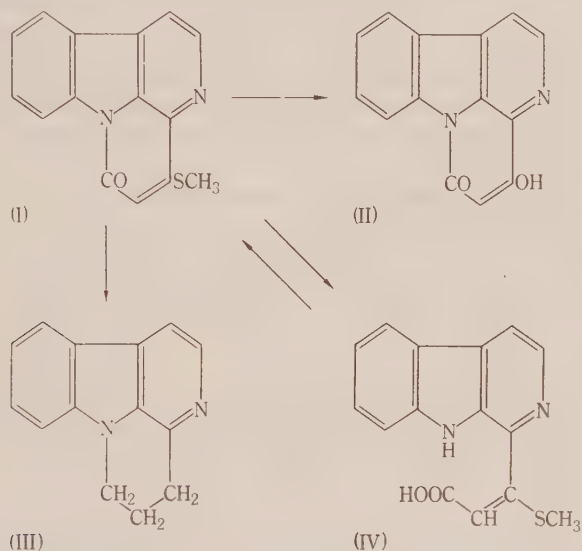


Fig. 2

evidently *cis*-2-methylthio-2 (1' β -carboxyl)acrylic acid (IV) as in aqueous alkali the lactam ring reforms and the alkaloid is regenerated. In the conversion of I to II, it is evidently IV which is formed first and from which the methylthio-group is eliminated, the product then undergoing recyclization to the resonance-stabilized 2,4-dihoxypyridine derivative II.

Confirmation of the presence of both a methylthio-group and of the canthine ring system is furnished by reduction of the alkaloid with zinc and hydrochloric acid which gives methyl mercaptan and 4,5-dihydrocanthine (III). The unexpected, smooth reduction of the lactam group in this class of compounds has already been reported by Haynes, Nelson, and Price (1952). Further support for the structure I is provided by the ultraviolet absorption spectrum (Fig. 2 and Table 1) which resembles closely the spectra of canthin-6-one and the 4- and 5-methoxycanthinones, and by the following reactions.

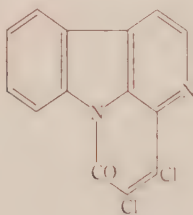


A benzene solution of the alkaloid with excess Raney nickel undergoes desulphurization together with hydrogenation producing 4,5-dihydrocanthinone; when an aged Raney catalyst was employed desulphurization took place unaccompanied by hydrogenation, the product being canthin-6-one. Spero, McIntosh, and Levin (1948) observed that the outcome of desulphurization of the thiol ester of a steroid acid also depended on the activity of the Raney nickel catalyst. Whereas freshly prepared catalyst gave the alcohol, a partially deactivated catalyst gave the aldehyde. If the alkaloid is reduced with Raney nickel in alcoholic solution, alcoholysis may also occur (cf. Haynes, Nelson, and Price 1952). Thus in methanol the methyl ester of 2-(1'- β -carbolyl)propionic acid is formed, whereas in ethanol the product is either 4,5-dihydrocanthinone or the ethyl ester of 2-(1'- β -carbolyl)propionic acid depending on the time of contact with the catalyst. Reduction of the alkaloid with tin and hydrochloric acid gives β -carbolylpropionic acid.

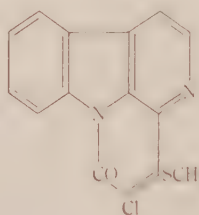
Finally, the structure I has been fully established by synthesis from 4-hydroxycanthin-6-one and therefore from β -carboline-1-carboxylic acid. 4-Chlorocanthinone was prepared from II by the action of phosphoryl chloride and this when heated in a sealed tube with potassium methyl mercaptide in alcohol readily gave a crystalline product identical with the alkaloid. In that it contains a methylthio-group the substance is unique among plant alkaloids.

II. REACTIONS OF 4-METHYLTHIOCANTHINONE AND SOME DERIVATIVES

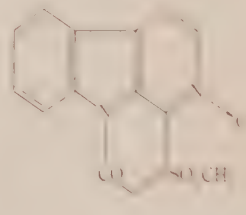
The conversion of 4-hydroxycanthinone to 4-chlorocanthinone by phosphoryl chloride parallels the preparation by Lutz *et al.* (1946) of 4,7-dichloro-1-methyl-2-quinolone from 7-chloro-4-hydroxy-1-methyl-2-quinolone by means of phosphoryl chloride. For the analogous conversion of 4-hydroxy-1-methyl-2-quinolone to 4-chloro-1-methyl-2-quinolone the same workers used a mixture of phosphoryl chloride and phosphorus pentachloride, but when 4-hydroxycanthinone is treated with such a mixture a dichloro-compound $C_{14}H_6ON_2Cl_2$ is formed, presumably 4,5-dichlorocanthin-6-one (V), together with an acidic substance giving an acetyl derivative $C_{16}H_9O_3N_2Cl$. The acidic substance, regarded as 4-hydroxy-5-chlorocanthinone, may be an intermediate in the formation of the dichloro-compound, or it may arise by hydrolysis during the working up of the reaction mixture. The dichlorocanthinone (V), treated with Raney nickel followed by alkali, gives 2-(1'- β -carbolyl)propionic acid, this constituting in effect a synthesis of canthin-6-one.



(V)



(VI)

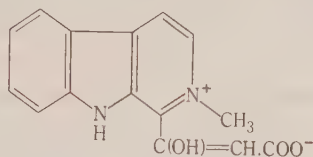


(VII)

Reaction of 4-methylthiocanthinone with "Chloramine T", carried out with the object of preparing a derivative of the sulphilimine type described by Mann and Pope (1922), resulted instead in chlorination. The product $C_{15}H_9OSN_2Cl$ is presumably 5-chloro-4-methylthiocanthinone (VI). Attempts at oxidizing the alkaloid to a sulphone by means of nitric acid or permanganate were unsuccessful, while hydrogen peroxide in acetic acid brought about the addition of three atoms of oxygen giving a product $C_{15}H_{10}O_4SN_2$ which is evidently the sulphone N oxide (VII). Like canthinone N oxide it is not reduced by sulphur dioxide (Haynes, Nelson, and Price 1952), but with zinc and acetic acid is converted to 4,5-dihydrocanthinone, cleavage of the $C-SO_2CH_3$ bond taking place as well as reduction of the N oxide. Compared with canthinone and 5-methoxycanthinone, 4-methylthio- and 4-methoxycanthinone are very resistant to oxidation by permanganate in acetone, presumably due to stabilization of the 4,5-double bond by resonance between the methylthio- (or methoxyl) group and the lactam $>C=O$ group.

The yield of 4-methoxycanthinone from the action of diazomethane on 4-acetoxycanthinone is only *c.* 35 per cent., the remainder of the starting material being converted to a red solid $C_{15}H_{12}O_3N_2$ soluble in water and alcohols but practically insoluble in non-hydroxylic solvents. The substance contains one methylimino-group but no methoxyl group and the structure VIII is suggested for it. 4-Methoxycanthinone, when heated for a few minutes with amyl alcoholic

alkali is converted by opening of the lactam ring to an acid. The formation of this acid, as shown by analysis of its *methyl ester* $C_{20}H_{22}O_3N_2$, involves an alkoxyl interchange, that is, the acid is 2-amyloxy-2-(1'- β -carbolyl)acrylic acid.



(VIII)

4-Methoxycanthinone is demethylated by hydrobromic acid in acetic acid and some 4-hydroxycanthinone is also formed as a by-product in the action with amyl alcoholic alkali.

III. THE BASICITY OF 4-METHYLTHIOCANTHINONE

The structure I established for the alkaloid accounts satisfactorily for the majority of its reactions and properties, but there are certain points, notably the basicity, which merit further discussion. Judged by solubility in aqueous acids, an approximate order of basic strengths of the *Pentaceras* alkaloids and their derivatives is as follows:

β -Carboline, 2-(1'- β -carbolyl)propionic acid esters, 4,5-dihydrocanthine, 4,5-dihydrocanthinone, *trans*-2-(1'- β -carbolyl)acrylic acid methyl ester, *cis*-2-methylthio-2-(1'- β -carbolyl)acrylic acid methyl ester > canthin-6-one, 5-methoxycanthin-6-one > 4-methoxycanthin-6-one > 4-chlorocanthin-6-one, 4-methylthiocanthin-6-one.

Thus 4,5-dihydrocanthinone is easily soluble in 0.01N hydrochloric acid, while 4-methylthiocanthinone is only slightly soluble in 2N though it dissolves readily in 4N hydrochloric acid, solution being followed by the separation of a yellow hydrochloride. Moreover, canthinone and 4- and 5-methoxycanthinones form picrates and methiodides, while 4-methylthiocanthinone forms neither a picrate nor a methiodide being recovered unchanged after heating with methyl iodide at 100 °C. Albert, Goldacre, and Phillips (1948) have pointed out the danger of drawing conclusions as to relative basic strengths from qualitative observations, nevertheless the differences encountered warrant some discussion even though this is limited by the qualitative nature of the observations. The similarity in basic strength of 4,5-dihydrocanthine, 4,5-dihydrocanthinone, and the esters of β -carbolylpropionic acid to β -carboline is to be expected, while withdrawal of electrons from N_{py} by the lactam $>C=O$ group should lead to canthin-6-one being somewhat more weakly basic than β -carboline. Likewise, the basicity of the chloro-compound in relation to that of β -carboline and canthinone is consistent with the inductive influence of the chlorine atom. The result of introducing a methoxyl group is not so easily predicted (cf. Albert and Goldacre 1946) but it is apparent that in 5-methoxycanthinone resonance and inductive effects more or less counterbalance, while in the 4-methoxy-compound the base-weakening inductive effect predominates or is supplemented by some

other influence, such as steric hindrance to the approach of the hydronium ion (cf. Albert and Goldacre 1944). The problem is essentially one of difference between the effects of a methoxyl and a methylthio-group, and the steric factor may be to some extent responsible at least as regards methiodide formation. A second factor which may possibly influence acid solubility is that the sulphur atom in sulphides has little, if any, ability to act as a donor in hydrogen bond formation (Tarbell and Harnish 1951). Thus, a proton, after addition to N_{py} , might be held more strongly in the 4-methoxy-compound as a result of hydrogen bonding with the methoxyl oxygen than in the 4-methylthio-compound in which hydrogen bonding would not occur. Hearn, Morton, and Simpson (1951) state that there appears to be no hydrogen bonding in 8-hydroxyquinoline as the spectra of the 8-hydroxy- and 8-methoxy-compounds do not differ from one another in neutral or in acid solution, and, though the cases are not strictly comparable, it would seem that the effect of such hydrogen bonding, if operative in 4-methoxycanthinone, would be small.

The relatively strongly basic character of the methyl esters of 2-amyloxy- and 2-methylthio-2-(1'- β -carbonyl)acrylic acids indicates that, on opening the lactam ring, interaction between the acrylic acid side-chain and the β -carboline ring system is considerably reduced. In agreement with this, both esters are colourless and the ultraviolet absorption spectrum of the methylthio-compound is very similar to that of β -carboline. On the other hand, methyl β -carbonyl-acrylate itself is yellow and the long-wave absorption maximum is displaced c. 30 m μ to longer wavelengths as compared with that of the methyl ester of 2-methylthio-2-(1'- β -carbonyl)acrylic acid. Presumably for steric reasons, there is a lesser degree of conjugation between side-chain and nucleus after opening the lactam ring in the presence of the amyloxy- or methylthio-substituent.

The ultraviolet absorption spectra of 4-methoxy- and 4-methylthiocanthinone are similar, the only result of the replacement of oxygen by sulphur being that the two longer wavelength band systems are displaced c. 10 m μ to longer wavelengths with no significant change in intensity. In the few other instances which have been recorded the effect on the absorption spectrum varies considerably. Passing from anisole to methylthiobenzene there is a displacement of the band 20 m μ to *shorter* wavelengths with a fivefold increase in extinction coefficient (Scheibe 1926; Fehnel and Carmack 1949). On the other hand, a displacement of a similar magnitude to *longer* wavelengths with greater intensity is observed on passing from 2-methoxyquinoline to 2-methylthio-4-methylquinoline (Ley and Specker 1939; Morton and Stubbs 1939), while comparison of the spectra of 1-methylthiobenzthiazole and 1-methoxy-5-methylbenzthiazole shows only a small wavelength but a large intensity difference (Hunter and Parken 1935; Morton and Stubbs 1939).

IV. EXPERIMENTAL

All melting points are corrected except where otherwise stated. Microanalyses were carried out in the C.S.I.R.O. Microanalytical Laboratory.

(a) *Properties of the Alkaloid C₁₅H₁₀OSN₂*

The alkaloid is sparingly soluble in ethanol but readily in chloroform and crystallizes from a chloroform-ethanol mixture as colourless needles, m.p. 252.5–253.5 °C. Analysis showed the absence of methoxyl and methylimino-groups; other analytical data have been reported previously (Haynes, Nelson, and Price 1952). A solution in chloroform (c, 2.0) was optically inactive.

The alkaloid is only slightly soluble in 2N aqueous hydrochloric acid but dissolves in 4N hydrochloric acid and a yellow hydrochloride separates on standing. It did not form a picrate and was recovered unchanged (m.p. and mixed m.p.) after attempted preparation of a methiodide by heating in a sealed tube with methyl iodide at 100 °C. for 6 hours. The alkaloid was also recovered (m.p. and mixed m.p.) after treatment with (i) boiling acetic anhydride containing a little pyridine for 2 hours, (ii) nitrous acid, (iii) ethanolic 2,4-dinitrophenylhydrazine sulphate, (iv) boiling ethanolic (2.5%) or aqueous (5%) hydrochloric acid for 16 hours, and (v) a mixture of hydrobromic acid (46% aqueous; 1 part) and acetic acid (2 parts) at the boiling point for 24 hours.

Owing to its inability to form water-soluble salts, the alkaloid has not been examined pharmacologically.

(b) *Reaction with Alcoholic Alkali*

(i) The alkaloid (5 g.), boiled under reflux with a solution of potassium hydroxide (25 g.) in ethanol (250 ml.), dissolved after several minutes, and the orange solution slowly deposited a yellow solid. The reaction mixture smelt strongly of mercaptan but only traces of mercaptide were precipitated when nitrogen was passed from the reaction vessel into 10% mercuric cyanide solution. Refluxing was discontinued after c. 40 hours when no further solid was separating and the mixture filtered. Examination of the filtrate gave only negligible amounts of acidic or basic material and a further attempt to identify mercaptan after acidification was unsuccessful.

The yellow precipitate was dissolved in water and the solution acidified giving 4-hydroxycanthin-6-one (3.0 g.) which separated from acetic acid as a yellow microcrystalline powder, m.p. 288–290 °C. (uncorr.).

Found: C, 70.9; H, 3.6; N, 11.7%.

Calculated for C₁₄H₈O₂N₂: C, 71.2; H, 3.4; N, 11.9%.

The substance dissolves in sodium bicarbonate, the sodium salt being only sparingly soluble in the presence of excess alkali; it separates as pale yellow plates from 10% sodium hydroxide. 4-Hydroxycanthinone was recovered unchanged after refluxing a solution in acetic acid with *o*-phenylenediamine for 1 hour, and after attempted reduction with Raney alloy and sodium hydroxide. It gives a red colour with ethanolic ferric chloride. Refluxing with acetic anhydride and pyridine for 15 minutes gave 4-acetoxycanthin-6-one, cream needles from chloroform-light petroleum, m.p. 205.5–206.5 °C.

Found: C, 69.4; H, 3.5; N, 10.1; CH₃CO, 15.0%.

Calculated for C₁₄H₇ON₂(OOC.CH₃): C, 69.1; H, 3.6; N, 10.1; CH₃CO, 15.5%.

Oxidation of 4-hydroxycanthinone in aqueous alkaline solution with permanganate gave β-carboline-1-carboxylic acid, yellow needles from acetic acid, m.p. 238–238.5 °C. (decomp.). Decarboxylation by heating at 240–250 °C. for 1 minute in dibutyl phthalate gave β-carboline, identified by comparison (m.p. and mixed m.p.) with synthetic specimens of the base and of its picrate. The identity of the acid was confirmed by preparation of the methyl ester (thionyl chloride followed by methanol), colourless needles from benzene, m.p. 165–166 °C., mixed m.p. with a synthetic specimen of methyl β-carboline-1-carboxylate, 166–167 °C.

Found: N, 12.2%.

Calculated for C₁₃H₁₀O₂N₂: N, 12.4%.

(ii) A solution of the alkaloid (1.0 g.) and potassium hydroxide (7 g.) in amyl alcohol (70 ml.) was refluxed for 8 hours. Nitrogen was bubbled through the reaction mixture, then through a 10% solution of mercuric cyanide, precipitating mercury methyl mercaptide in poor yield.

After washing and recrystallization from ethanol it had m.p. and mixed m.p. 175 °C. 2,4-Dinitro-phenyl methyl sulphide was also prepared by washing the effluent nitrogen with ethanolic sodium hydroxide and adding to this an ethanolic solution of 2,4-dinitrochlorobenzene. It formed as a yellow precipitate, m.p. 127.5–128 °C. after two recrystallizations from methanol and showed no m.p. depression when mixed with an authentic specimen. The yield was again poor.

Found: S, 14.8%.

Calculated for $C_7H_6O_4SN_2$: S, 15.0%.

4-Hydroxycanthinone was recovered from the reaction mixture in 70% yield as the amyl alcohol-insoluble potassium salt.

(iii) The reaction with amyl alcoholic potassium hydroxide was carried out as described in (ii) above but, after refluxing for only 15 minutes, the solution was filtered, cooled in ice, diluted with ether, and extracted with ice-water. The aqueous extracts after washing with ether were cautiously acidified with ice-cold 5% acetic acid until a slimy yellow precipitate appeared. This was removed, the solution overlaid with ether, and acidification continued. The resulting white precipitate, twice crystallized from aqueous methanol and once from methanol-ether, gave colourless prisms melting with effervescence at 240 °C. (uncorr.) when placed in a bath preheated to c. 235 °C. Starting from lower temperatures it sintered but did not give a clear melt until 251–252 °C.

Found: C, 63.1; H, 4.1; N, 9.8; S, 11.2%.

Calculated for $C_{15}H_{12}O_2SN_2$: C, 63.4; H, 4.2; N, 9.9; S, 11.3%.

The acid was readily soluble in 0.1N hydrochloric acid and in cold aqueous sodium hydroxide but the alkaloid (m.p. and mixed m.p.) separated quantitatively from the alkaline solution at a rate dependent on the temperature. With methanolic hydrogen chloride it gave the *methyl ester*, colourless needles from benzene-light petroleum, m.p. 181–182 °C.

Found: C, 64.5; H, 4.6; N, 9.2; S, 10.9; CH_3O , 10.6%.

Calculated for $C_{16}H_{14}O_2SN_2$: C, 64.4; H, 4.7; N, 9.4; S, 10.7; CH_3O , 10.4%.

The ester was easily soluble in 0.03N hydrochloric acid. It was found to have no noteworthy physiological effect on mice or toads.

(c) *Synthesis of 4-Hydroxycanthin-6-one (II)*

β -Carboline-1-carboxylic acid (finely powdered, 1.0 g.) was refluxed with freshly distilled thionyl chloride (25 ml.) for 10 minutes.* The hydrochloride of the acid chloride, after distilling off excess thionyl chloride and drying in a vacuum over potassium hydroxide, was powdered and added gradually with shaking to the magnesium ethoxy-derivative of malonic ester, prepared by the procedure of Walker and Hauser (1946) from 0.3 g. magnesium, and diluted with a further 30 ml. anhydrous ether. The mixture was refluxed for 15 minutes, the ethereal solution decanted from unreacted greenish solid and acidified with acetic acid giving a pale green precipitate. This was refluxed for 4 hours with a mixture of concentrated hydrochloric acid (7.5 ml.) and ethanol (7.5 ml.). The yellow solid which separated on cooling was dissolved in dilute sodium hydroxide, reprecipitated by acetic acid, and thrice crystallized from acetic acid giving yellow needles, m.p. 288–290 °C. (uncorr.) undepressed by admixture with a specimen prepared from the alkaloid. Yield 0.15 g.

Found: C, 70.9; H, 3.6; N, 11.8%.

Calculated for $C_{14}H_8O_2N_2$: C, 71.2; H, 3.4; N, 11.9%.

4-Acetoxycanthinone, colourless needles from chloroform-light petroleum, melted at 206–206.5 °C. alone or mixed with a specimen obtained from the alkaloid.

Found: C, 69.0; H, 3.5; N, 10.1; CH_3CO , 15.6%.

Calculated for $C_{16}H_{10}O_3N_2$: C, 69.1; H, 3.6; N, 10.1; CH_3CO , 15.5%.

* No attempt was made to purify or analyse the acid chloride, but when treated with methanol, methyl β -carboline-1-carboxylate was produced in good yield.

(d) 4-Methoxycanthin-6-one

A mixture of 4-methoxycanthinone (0.5 g.) in benzene (10 ml.) and of benzene (from 3 g.) of concentrated sulphuric acid was stirred and after 3 days the red crystalline solid formed was filtered off. The filtrate was evaporated, added to excess sodium hydroxide solution (2%), in the cold, and the 4-methoxycanthinone extracted with chloroform and purified by chromatographing a benzene solution on alumina. It formed colourless needles from benzene, m.p. 220-220.5 °C. Yield 0.3 g.

Found: C, 72.0; H, 4.0; N, 11.1; CH_2O , 12.4%.

Calculated for $\text{C}_{11}\text{H}_{10}\text{O}_2\text{N}_2$: C, 72.0; H, 4.0; N, 11.2; CH_2O , 12.4%.

The red crystals of acetone, representing the bulk of the starting material, was recrystallized once from ethanol, and then from water and obtained as red needles which had no sharp m.p. but decomposed at c. 240 °C.

Found: C, 67.0; H, 4.2; N, 10.5; CH_2N , 10.6%; CH_2O , nil.

Calculated for $\text{C}_{11}\text{H}_{12}\text{O}_2\text{N}_2$: C, 67.2; H, 4.5; N, 10.4; CH_2N , 10.8%.

The substance is freely soluble in hot water or ethanol, but insoluble in benzene, chloroform, or ether.

4-Methoxycanthinone is fairly sparingly soluble in 1 N but dissolves easily in 1 N hydrochloric acid and the solution slowly depositing pale yellow crystals of the hydrochloride. 4-Methoxycanthinone picrate separated from ethanol as yellow needles, m.p. 221.5-222.5 °C. (decomp.).

Found: N, 14.7%.

Calculated for $\text{C}_{11}\text{H}_{10}\text{O}_2\text{N}_2 \cdot \text{C}_6\text{H}_2\text{O}_7\text{N}_2$: N, 14.6%.

4-Methoxycanthinone methiodide, prepared by refluxing a solution of 4-methoxycanthinone in excess methanol for 7 hours, was obtained as orange needles, m.p. 218 °C. (decomp.) after crystallization from water containing a little hydriodic acid.

Found: C, 46.8; H, 3.7%.

Calculated for $\text{C}_{11}\text{H}_{12}\text{O}_2\text{N}_2 \cdot \text{HI} \cdot \text{H}_2\text{O}$: C, 46.8; H, 3.7%.

4-Methoxycanthinone was recovered unchanged after refluxing with a mixture of equal volumes of concentrated hydrochloric acid and ethanol for 2 hours, but was demethylated by refluxing for 3 hours with a mixture of equal volumes of 40% aqueous hydrobromic acid and ethanol. 4-Methoxycanthinone being identified by its m.p. and mixed m.p. and by conversion to 4-acetoxycanthinone.

(iii) A solution of 4-methoxycanthinone (250 mg.) in amyl alcoholic potassium hydroxide (10%, 20 ml.) was boiled for 7 minutes, cooled in ice, and filtered from some potassium salt of 4-methoxycanthinone (30 mg., identified as m.p. and mixed m.p. of 4-methoxycanthinone and its acetyl derivative). The filtrate was diluted with ice-cold chloroform (50 ml.), extracted with several 10 ml. portions of water, and the extracts combined with 50% acetic acid. The earlier extracts and the acid precipitated from them were yellow. Later extracts were colourless and gave no colour on acidification. After standing overnight at 5 °C. the acid (215 mg.) was filtered off and without further purification, identified by means of methanolic hydrogen chloride. The methanolic solution in benzene was chromatographed on alumina, and the main fraction (140 mg.) crystallized from 90% petroleum and obtained as colourless needles, m.p. 145-146 °C.

Found: C, 70.8; H, 6.5%.

Calculated for $\text{C}_{20}\text{H}_{22}\text{O}_2\text{N}_2$: C, 71.0; H, 6.5%.

The ester is soluble in 0.01N hydrochloric acid.

(e) Oxidation of 4-Methylthiocanthinone with Permanganate

To a boiling solution of the alkaloid (0.7 g.) in acetone (300 ml.) powdered potassium permanganate was added during 24 hours and the precipitated manganese dioxide filtered off together with some unchanged 4-methylthiocanthinone which separated on cooling. Evaporation of the filtrate and crystallization of the residue from chloroform-ethanol gave more 4-methylthiocanthinone, a total of 0.6 g. alkaloid being recovered. Treatment of the manganese dioxide

with sulphur dioxide gave β -carboline-1-carboxylic acid. The acid was purified through the sodium salt; yield 0.05 g., m.p. 238–238.5 °C. (decomp.), methyl ester, m.p. and mixed m.p. with an authentic specimen 165.5–166 °C.

By contrast with this low rate of oxidation, the oxidation of a 0.33% solution of canthinone in acetone was complete within 2 hours at room temperature and under the same conditions the oxidation of 5-methoxycanthinone required c. 18 hours. 4-Methoxycanthinone, however, resembled the 4-methylthio-compound in that in 18 hours only a negligible amount was oxidized by permanganate at room temperature.

(f) *Action of Zinc and Hydrochloric Acid on 4-Methylthiocanthinone*

A stream of nitrogen was passed through a suspension of the alkaloid (1.0 g.) in boiling aqueous hydrochloric acid (80 ml.), zinc dust added in several portions, and the issuing gases passed into 10% aqueous mercuric cyanide precipitating mercury mercaptide, m.p. and mixed m.p. 174–175 °C. after several crystallizations from ethanol.

Found: S, 21.9%.

Calculated for $\text{Hg}(\text{SCH}_3)_2$: S, 21.7%.

When the issuing gases were passed into ethanolic sodium hydroxide and 2,4-dinitrochlorobenzene added, 2,4-dinitrophenyl methyl sulphide, m.p. and mixed m.p. 127.5–128 °C., was formed.

When the reaction was complete, the hydrochloric acid solution was filtered, cooled, and the resulting colourless crystalline zinc complex dissolved in water and basified with excess sodium hydroxide. Extraction with chloroform yielded an oily base which was dissolved in benzene and the solution run through a column of alumina. The eluted base separated from aqueous methanol as colourless plates, m.p. 70–72 °C. not depressed by admixture with dihydroecanthine hydrate prepared from canthinone (cf. Haynes, Nelson, and Price 1952).

Found: C, 74.0; H, 6.3; N, 12.3%.

Calculated for $\text{C}_{14}\text{H}_{12}\text{N}_2 \cdot \text{H}_2\text{O}$: C, 74.3; H, 6.2; N, 12.4%.

The picrate separated from ethanol as yellow needles, m.p. 258 °C. (decomp.) after previous darkening, this behaviour being unaltered by admixture with authentic dihydroecanthine picrate.

Found: C, 55.1; H, 3.6; N, 15.8%.

Calculated for $\text{C}_{14}\text{H}_{12}\text{N}_2 \cdot \text{C}_6\text{H}_3\text{O}_7\text{N}_3$: C, 54.9; H, 3.4; N, 16.0%.

(g) *Action of Raney Nickel on 4-Methylthiocanthinone*

(i) A solution of the alkaloid (0.5 g.) in benzene (75 ml.) was refluxed with Raney nickel (c. 4 g.) for 3 hours, the nickel filtered off, washed, and the combined filtrate and washings evaporated to dryness. The residue (0.1 g.) crystallized from benzene-light petroleum as colourless leaflets, m.p. 128.5–130 °C. undepressed by admixture with 4,5-dihydroecanthinone. This result was obtained with several freshly prepared batches of Raney nickel, but in two experiments with an aged sample of the catalyst, the product was canthinone (0.2 g. from 0.5 g. 4-methylthiocanthinone), m.p. and mixed m.p. 160–161 °C.

(ii) A solution of 4-methylthiocanthinone (1.0 g.) in methanol (150 ml.) was refluxed with Raney nickel (c. 10 g.) for 4 hours, filtered, and the combined filtrate and washings from the nickel evaporated to dryness. The residue (0.3 g.) crystallized from benzene-light petroleum as colourless needles, m.p. 145.5–146 °C. undepressed by admixture with the methyl ester of 2-(1'- β -carbonyl)propionic acid.

(iii) 4-Methylthiocanthinone (0.5 g.) in ethanol (80 ml.) refluxed with Raney nickel (c. 5 g.) for 4 hours, then filtered immediately, gave 4,5-dihydroecanthinone, m.p. and mixed m.p. 129–130 °C. after repeated crystallization from light petroleum. When the reaction mixture stood overnight in contact with the nickel the product was largely the ethyl ester of β -carbonyl-propionic acid, obtained, after repeatedly crystallizing from aqueous methanol and finally from light petroleum as colourless needles, m.p. 126–127 °C. undepressed by admixture with an authentic specimen.

Found: C, 71.8; H, 5.7; N, 10.7; $\text{C}_2\text{H}_5\text{O}$, 17.3%.

Calculated for $\text{C}_{16}\text{H}_{16}\text{O}_2\text{N}_2$: C, 71.6; H, 6.0; N, 10.4; $\text{C}_2\text{H}_5\text{O}$, 16.8%.

(h) Action of Tin and Hydrochloric Acid on 4-Methylthiocanthinone

The alkaloid (1.0 g.) in hydrochloric acid (5%, 150 ml.) was boiled under reflux for 6 hours with granulated tin and the solution filtered and allowed to stand overnight. The colourless solid which separated was filtered off, dissolved in hot hydrochloric acid (1%), and the solution cooled and saturated with hydrogen sulphide. After filtration from stannous sulphide, β -carbolyl-propionic acid was precipitated by the addition of sodium bicarbonate. The acid was purified by passage through the sodium salt and esterified with methanolic hydrogen chloride giving the methyl ester, m.p. and mixed m.p. 145–146 °C.

(i) Attempts to Prepare a Sulphone from the Alkaloid

4-Methylthiocanthinone was recovered unchanged (m.p. and mixed m.p.) after boiling for 1 minute with 68% nitric acid. The dropwise addition of aqueous potassium permanganate (0.1N, 100% excess) to a cooled solution in acetic acid also led to the recovery of the greater part of the alkaloid.

A solution of 4-methylthiocanthinone (0.65 g.) in acetic acid (75 ml.) containing hydrogen peroxide (30%, 1 ml.), allowed to stand at room temperature for 9 days, deposited the *sulphone-N-oxide* (0.6 g.) which crystallized from chloroform (in which it is only sparingly soluble) as cream prisms, m.p. 201 °C. (decomp.).

Found: C, 57.8; H, 3.3; N, 8.7; S, 10.4%.

Calculated for $C_{15}H_{10}O_4SN_2$: C, 57.3; H, 3.2; N, 8.9; S, 10.2%.

From acetic acid the substance separated as orange-yellow plates melting with decomposition at 205 °C.; analytical results were the same as for the cream modification. When a solution in acetic acid was treated with sulphur dioxide and allowed to stand overnight the substance was recovered unchanged. Reduction by zinc in boiling acetic acid solution gave dihydro-canthinone, colourless needles from light petroleum, m.p. and mixed m.p. 128–129 °C.

(j) Reaction of 4-Methylthiocanthinone with "Chloramine T"

A solution of 4-methylthiocanthinone (1.0 g.) and "Chloramine T" (2.5 g.) in acetic acid (15 ml.) was heated for 2 hours on the water-bath. The product, which contained chlorine, crystallized from chloroform-ethanol as pale yellow microscopic prisms (0.4 g.) which melted at 252–253 °C. (uncorr., decomp.) after softening and darkening from 239 °C.

Found: C, 59.8; H, 3.3; S, 10.4%.

Calculated for $C_{15}H_9OSN_2Cl$: C, 59.9; H, 3.0; S, 10.6%.

(k) Action of Phosphorus Halides on 4-Hydroxycanthinone

(i) 4-Hydroxycanthinone (0.6 g.) was refluxed with freshly distilled phosphorus oxychloride (30 ml.) for 3 hours, the latter distilled off under reduced pressure, and the residue treated with cold sodium hydroxide solution and extracted with chloroform. The crude chloro-compound remaining after evaporation of the chloroform was dissolved in benzene, chromatographed on alumina, and the middle fractions combined and crystallized from chloroform-ethanol. 4-Chloro-canthinone (0.31 g.) separated as cream needles, m.p. 201–202 °C.

Found: C, 66.1; H, 3.1%.

Calculated for $C_{14}H_7ON_2Cl$: C, 66.0; H, 2.8%.

The substance is only sparingly soluble in 1N but dissolves readily in 3N hydrochloric acid.

(ii) 4-Hydroxycanthinone (1.0 g.), phosphorus pentachloride (5 g.), and phosphorus oxychloride (1 ml.) were heated together for 3 hours at 130–140 °C., the reaction mixture cooled, poured into ice-water, and neutralized with sodium hydroxide when a yellow precipitate formed. The whole was extracted repeatedly with chloroform, after which the yellow precipitate was filtered off, dissolved in hot aqueous sodium hydroxide, and the solution filtered and acidified with acetic acid giving a brick-red precipitate (0.45 g.). The substance, which did not melt

but decomposed gradually up to 360 °C., was boiled for 10 minutes with acetic anhydride containing a trace of pyridine, giving an *acetyl derivative*, colourless needles from benzene, m.p. 225–226 °C.

Found: C, 61.3; H, 2.8; N, 9.1; Cl, 11.4%.

Calculated for $C_{16}H_9O_3N_2Cl$: C, 61.4; H, 2.9; N, 9.0; Cl, 11.4%.

The chloroform extract was evaporated, the residue dissolved in benzene and chromatographed on alumina, and the middle fractions, melting between 216 and 220 °C., combined and crystallized from chloroform-ethanol. 4,5-Dichlorocanthinone separated as pale yellow needles, m.p. 226–227 °C. Yield 0.3 g.

Found: C, 58.2; H, 2.3; N, 9.9; Cl, 24.3%.

Calculated for $C_{14}H_6ON_2Cl_2$: C, 58.1; H, 2.1; N, 9.7; Cl, 24.6%.

The substance is only slightly soluble in 2N but dissolves in 3N hydrochloric acid, the solution depositing a yellow crystalline hydrochloride.

(l) *Synthesis of 2-(1'-β-Carbonyl)propionic Acid*

A solution of 4,5-dichlorocanthinone (100 mg.) in methanol (60 ml.) was boiled under reflux for 4 hours with Raney nickel (c. 1 g.), filtered, the nickel washed with methanol and the combined filtrate and washings evaporated to dryness. The residue was extracted with hot 10% sodium hydroxide solution and the extract filtered and cooled when a colourless crystalline sodium salt separated. The sodium salt was filtered off, dissolved in water, and the solution acidified with acetic acid, giving β-carbonylpropionic acid, identified by conversion to its methyl ester, m.p. and mixed m.p. 145–146 °C.

Found: C, 70.9; H, 5.3%.

Calculated for $C_{15}H_{14}O_2N_2$: C, 70.9; H, 5.5%.

(m) *Synthesis of 4-Methylthiocanthin-6-one*

A solution of potassium methyl mercaptide, prepared by bubbling excess methyl mercaptan into ethanol (5 ml.) containing potassium hydroxide (50 mg.), was added to 4-chlorocanthinone (156 mg.) in ethanol (15 ml.) and the mixture heated in a sealed tube at 75–80 °C. for 7 hours. Crystalline material which had separated was filtered off, the filtrate made just acid with acetic acid and evaporated to a small volume. Dilution with water and extraction with chloroform gave a further quantity of crystalline solid. The combined solids were dissolved in benzene and chromatographed on alumina, taking 5 fractions as follows:

- (1) Gum.
- (2) White crystalline solid, m.p. 248–250 °C.
- (3) White crystalline solid, m.p. 248–250 °C.
- (4) White crystalline solid, m.p. 204–210 °C.
- (5) White crystalline solid, m.p. 190–192 °C.

Fractions (2) and (3) combined (45 mg.) were crystallized from chloroform-ethanol, giving 4-methylthiocanthinone, almost colourless, fine needles, m.p. 250–252 °C., mixed m.p. with the alkaloid 250–252 °C.

Found: C, 68.0; H, 3.9; S, 11.8%.

Calculated for $C_{15}H_{10}OSN_2$: C, 67.7; H, 3.8; S, 12.0%.

The identity was confirmed by running paper chromatograms of the natural and synthetic alkaloid with butanol-acetic acid at 17 °C., the mobile phase being prepared by equilibrating butanol with 5% aqueous acetic acid. The R_F values were as follows:

Natural 4-methylthiocanthinone	0.92,
Synthetic 4-methylthiocanthinone	0.93.

(n) *Ultraviolet Absorption Spectra*

The ultraviolet absorption spectra were measured in dioxan solution by means of a Beckmann model DU spectrophotometer.

V. ACKNOWLEDGMENTS

The authors are indebted to Professor E. J. Hartung for providing accommodation and facilities in the Chemistry Department, University of Melbourne; to Associate Professor F. H. Shaw for carrying out pharmacological tests; and to Dr. H. H. Hatt for helpful criticism of the manuscript.

VI. REFERENCES

- ALBERT, A., and GOLDACRE, R. (1944).—*Nature* **153**: 467.
ALBERT, A., and GOLDACRE, R. (1946).—*J. Chem. Soc.* **1946**: 706.
ALBERT, A., GOLDACRE, R., and PHILLIPS, J. (1948).—*J. Chem. Soc.* **1948**: 2240.
FEHNEL, E. A., and CARMACK, M. (1949).—*J. Amer. Chem. Soc.* **71**: 84.
HAYNES, H. F., NELSON, E. R., and PRICE, J. R. (1952).—*Aust. J. Sci. Res. A* **5**: 387.
HEARN, J. M., MORTON, R. A., and SIMPSON, J. C. E. (1951).—*J. Chem. Soc.* **1951**: 3319.
HORNER, L. (1939).—*Liebigs Ann.* **540**: 73.
HUNTER, R. F., and PARKEN, E. R. (1935).—*J. Chem. Soc.* **1935**: 1755.
LEY, H., and SPECKER, H. (1939).—*Ber. dtsh. chem. Ges.* **72**: 192.
LUTZ, R. E., CODINGTON, J. F., ROWLETT, R. J., DEINET, A. J., and BAILEY, P. S. (1946).—*J. Amer. Chem. Soc.* **68**: 1810.
MANN, F. G., and POPE, W. J. (1922).—*J. Chem. Soc.* **121**: 1052.
MORTON, R. A., and STUBBS, A. L. (1939).—*J. Chem. Soc.* **1939**: 1321.
NELSON, E. R., and PRICE, J. R. (1952).—*Aust. J. Sci. Res. A* **5**: 563.
SCHEIBE, G. (1926).—*Ber. dtsh. chem. Ges.* **59**: 2617.
SNYDER, H. R., WALKER, H. G., and WERBER, F. X. (1949).—*J. Amer. Chem. Soc.* **71**: 527.
SPERO, G. B., MCINTOSH, A. V., and LEVIN, R. H. (1948).—*J. Amer. Chem. Soc.* **70**: 1907.
TARBELL, D. S., and HARNISH, D. P. (1951).—*Chem. Rev.* **49**: 1.
WALKER, H. G., and HAUSER, C. R. (1946).—*J. Amer. Chem. Soc.* **68**: 1386.

SHORT COMMUNICATIONS

COSMIC RAY ASYMMETRY STUDIES AT HOBART, TASMANIA*

By D. W. P. BURBURY† and A. C. McLAREN‡

The apparatus used previously by Burbury and Fenton (1952) has been modified by the introduction of a third counter tray and used with 12 cm. of lead absorber to obtain 91 days' measurements of the east-west asymmetry at a zenith angle of 45° and 56 days' measurements of the north-south asymmetry at 30° .

For correlation of the asymmetry with the radiosonde data, it was necessary to choose a period over which the radiosonde data could be assumed to represent the state of the atmosphere fairly well, at the same time including as much of the cosmic ray data as possible. A 4 hr. period centred on the hour nearest the time the radiosonde balloon reached the 300 mb. level was chosen, and the mean asymmetry and mean barometric pressure calculated for this period.

The correlation was tried in the same way as used by Duperier (1949) for the total radiation. The asymmetry, the barometric pressure, the height of an isobaric layer, and the mean temperature over a region above the layer were used as variables for several layers in turn. None of the correlation coefficients computed reached more than the 10 per cent. level of significance, so it was concluded that the effect of these variables on the asymmetry is probably too small to be detected under the present experimental conditions.

With the possibility of a null result in mind, each set of asymmetry values was tested for homogeneity by analysis of variance methods (Tables 1 and 2) and, in each case, the estimate of the variance from between days was not found to be significantly greater than the estimate of the variance from within days.

We may take the null result of the analyses of variance to indicate that the asymmetries are not dependent upon factors which change from day to day to an extent which can be detected by the present apparatus. It follows that the east and west or north and south cosmic ray intensities are similarly affected by external conditions.

We may, therefore, compare their sum with the barometric pressure in the usual way and find the barometer coefficients for the radiation from the inclined directions. From the north-south run at 30° , we find a barometer coefficient of -1.90 ± 0.08 per cent. per cm. Hg and from the east-west run at 45° , we find a coefficient of -3.56 ± 0.05 per cent. per cm. Hg.

* Manuscript received July 23, 1952.

† C.S.I.R.O., attached to Physics Department, University of Tasmania.

‡ Physics Department, University of Tasmania.

A tendency for the barometer coefficient to increase with zenith angle has also been found by Rossi (1939) and Trumpy and Orlin (1941). This result can be expected since the change in grams per square centimetre of air along

TABLE 1

RESULTS OF EAST-WEST ASYMMETRY MEASUREMENTS

Zenith angle 45° , 12 cm. lead absorber. Analysis of variance of 91 days' results classified according to days. (Asymmetry $+0.03$) $\times 10^4$ as variable

Source	Sum of Squares	Degrees of Freedom	Mean Square
Within days	300925193	1001	300625
Between daily means ..	27593391	90	306593
Total	328518584	1091	301117

Variance ratio 1.01 not significantly greater than 1.

Mean E.-W. asymmetry (for this experiment only) $= 0.0177 \pm 0.0011$ (P.E.).

the paths of particles at a zenith angle θ accompanying a change in pressure is proportional to $\sec \theta$. Lindholm (1950) has found that the consistency of the observed barometer coefficient is increased when the effect of the upper-air

TABLE 2

RESULTS OF NORTH-SOUTH ASYMMETRY MEASUREMENTS

Zenith angle 30° , 12 cm. lead absorber. Analysis of variance of 56 days' results classified according to days. (Asymmetry $+0.03$) $\times 10^4$ as variable

Source	Sum of Squares	Degrees of Freedom	Mean Square
Within days	146643912	616	238058
Between days	15397135	55	279948
Total	162041047	671	241492

Variance ratio 1.175 not significantly greater than 1.

Mean N.-S. asymmetry $= 0.0056 \pm 0.0013$ (P.E.).

conditions is removed. However, we consider the actual numerical values obtained to be less important than the trend with zenith angle.

The results of the east-west run combined with earlier results give a value of the asymmetry of 0.0195 ± 0.0010 compared with the predicted value of 0.0192 (Burbury and Fenton 1952). The north-south asymmetry was found to be 0.0056 ± 0.0013 , which is different from zero, the north intensity being the greater (Burbury 1952). This agrees with a suggestion by Rose (1951) that in high latitudes the north-south asymmetry is almost as pronounced as the east-west asymmetry.

The authors wish to thank Professor E. J. G. Pitman for his advice on the analysis of results, and the Commonwealth Meteorological Bureau for supplying the meteorological data.

References

- BURBURY, D. W. P. (1952).—*Phys. Rev.* **86**: 577.
BURBURY, D. W. P., and FENTON, K. B. (1952).—*Aust. J. Sci. Res. A* **5**: 47.
DUPERIER, A. (1949).—*Proc. Phys. Soc. A* **62**: 684.
LINDHOLM, F. (1950).—*Tellus* **2**: 63.
ROSE, D. C. (1951).—*Nature* **168**: 271.
ROSSI, B. (1939).—*Rev. Mod. Phys.* **11**: 281.
TRUMPY, B., and ORLIN, J. (1941).—*Arch. Math. Naturv.* **B 44**: 99.

ATTEMPTS TO EFFECT EXCHANGE AND DISPLACEMENT REACTIONS WITH COBALT(II) MESOPORPHYRIN*

By N. ASHELFORD† and D. P. MELLOR†

There is ample evidence that divalent metal atoms, such as those of Zn, Cu, Ni, Co, Fe, and Mn, can be introduced directly into the centre of the tetrapyrrolic ring system of many porphin derivatives and that under appropriate conditions these metal atoms can be equally readily removed from metal porphin complexes. In this respect metal porphin complexes differ from those of phthalocyanine from which it is extremely difficult if not impossible to remove these metal atoms without destruction of the molecule. To introduce metal atoms into phthalocyanine it is necessary to do so during the synthesis of the phthalocyanine ring system itself. In these circumstances no exchange of metal ions between metal phthalocyanine complexes can be expected. The object of this note is to describe experiments made in the attempt to discover whether cobalt ion can exchange with cobalt(II) mesoporphyrin and whether cobalt can be displaced from cobalt(II) mesoporphyrin by copper ion which would be expected to form a more stable complex with mesoporphyrin.

The exchange experiments were carried out in carefully dried pyridine,‡ one of the few solvents in which mesoporphyrin was sufficiently soluble for this work. Equimolar solutions of cobalt(II) mesoporphyrin and cobalt(II) acetate were made up in pyridine and equal volumes (25 ml.) of each were mixed together and allowed to stand in a thermostat at 25°C. for varying periods of time. Throughout the period of possible exchange, a stream of dry nitrogen

* Manuscript received July 28, 1952.

† Chemistry School, University of Sydney.

‡ In this solvent the exchange would, if it occurred, take place between $\text{Co}(\text{py})_4^{2+}$ and the dipyridine cobalt(II) mesoporphyrin.

was bubbled through the solution to prevent aerial oxidation. In one set of experiments ^{60}Co was incorporated in the mesoporphyrin and, in the other, with the cobalt(II) acetate.

After the lapse of the period of time decided upon for the exchange, the separation of cobalt(II) mesoporphyrin and cobalt(II) acetate was carried out as follows. To the pyridine solution were added 50 ml. of water and 50 ml. of chloroform. After thoroughly shaking the mixture and allowing it to stand 5 minutes, two layers separated; the lower, pyridine-chloroform layer containing cobalt(II) mesoporphyrin was then run off. The remaining water-pyridine layer containing the cobalt(II) acetate was washed four times with 25 ml. of chloroform each time. The combined chloroform-pyridine extractions were then similarly washed with 25 ml. of water and the washings added to the water-pyridine layer. The two soluble chemical compounds thus separated were then precipitated to determine their radioactivity. The cobalt(II) acetate was precipitated as cobalt(II) dipyridine thiocyanate by means of ammonium thiocyanate. Because of the relative difficulty of removing cobalt from cobalt(II) mesoporphyrin, its chloroform-pyridine solution was evaporated almost to dryness and a little water added. The remaining chloroform was boiled off leaving cobalt(III) mesoporphyrin suspended in water. The precipitate was filtered off and washed with water. A blank experiment with cobalt(II) acetate showed that 98–99 per cent. of the substance was extracted in the water-pyridine solution. With ^{60}Co in the mesoporphyrin, there was within the limits of experimental error, even over a long period of time, practically no exchange (Table 1).

TABLE 1
EXCHANGE OF ^{60}Co MESOPORPHYRIN WITH Co^{++} AT 20 °C.

Time (hr.)	Activity of Co^{II} Mesoporphyrin	Activity of Co^{II} Acetate	Exchange (%)
0	11241	62	1
1	8216	67	1
$3\frac{1}{2}$	7142	65	1
24	11061	108	1

Activities are given in counts per minute (corrected for background). When the exchange conditions were reversed, with ^{60}Co present as cobalt(II) acetate, there appeared to be an exchange that reached about 20 per cent. instantaneously and did not increase thereafter. This spurious result may have been due to the occlusion of free mesoporphyrin in the cobalt(II) mesoporphyrin in spite of every effort to prevent this. A more likely explanation of the effect is that cobalt(II) ion combined with the two adjacent propionic acid groups on the mesoporphyrin molecule, a probable reaction, especially in alkaline solution.

Attempts to carry out an exchange in glacial acetic acid solution failed owing to the low solubility of cobalt(II) mesoporphyrin in that solvent. Experiments in which pyridine was used as solvent were made to see whether copper

could replace cobalt in cobalt(II) mesoporphyrin. Even after 100 hours at 25 °C. the displacement amounted to not more than 10 per cent. (Table 2).

TABLE 2
REPLACEMENT OF CO FROM Co^{II} MESOPORPHYRIN BY CU

Time (hr.)	Activity of Co^{II} Mesoporphyrin	Activity of Co^{II} Acetate	Replacement (%)
0	6742	60	1
24	8417	134	2
100	7922	689	9

In these experiments the copper concentration greatly exceeded that of the cobalt compound, the concentration of which was approximately 10^{-5}M . Attempts to detect the replacement of cobalt with copper spectroscopically were unsuccessful. Barnes and Dorough (1950) were able completely to displace zinc from α , β , γ , δ tetraphenylporphin by copper only after 48 hours at the boiling point of pyridine.

From these experiments and those of other authors (Ruben *et al.* 1942) on similar compounds it must be concluded that metal atoms, such as Cu, Co, and Ni, are very firmly bound to the porphin nucleus being neither readily exchanged nor displaced by other metal atoms.*

References

- BARNES, J. W., and DOROUGH, G. D. (1950).—*J. Amer. Chem. Soc.* **72**: 4045.
RUBEN, S., KAMEN, M. D., ALLEN, M. B., and NAHINSKY, P. (1942).—*J. Amer. Chem. Soc.* **64**: 2297.

ON *N*-(2,4-DINITROPHENYL)PHTHALIMIDE†

By A. L. BECKWITH‡ and J. MILLER‡

When sodium phthalimide is reacted with 1-chloro-2,4-dinitrobenzene in methanol or water, the products are 2,4-dinitroanisole and 2,4-dinitrophenol respectively. The reactions described below show that these substances are formed by direct solvolysis and not via *N*-(2,4-dinitrophenyl)phthalimide. In boiling acetone no reaction occurs but *N*-(2,4-dinitrophenyl)phthalimide can be prepared by melting the two compounds together at 200–210 °C. for 2 hours

* This statement does not hold for porphin complexes of alkali and alkaline earth metals. Neither does it hold for Cu, Co, and Ni complexes of a quadridentate chelate molecule like salicylaldehyde ethylenediamine.

† Manuscript received July 7, 1952.

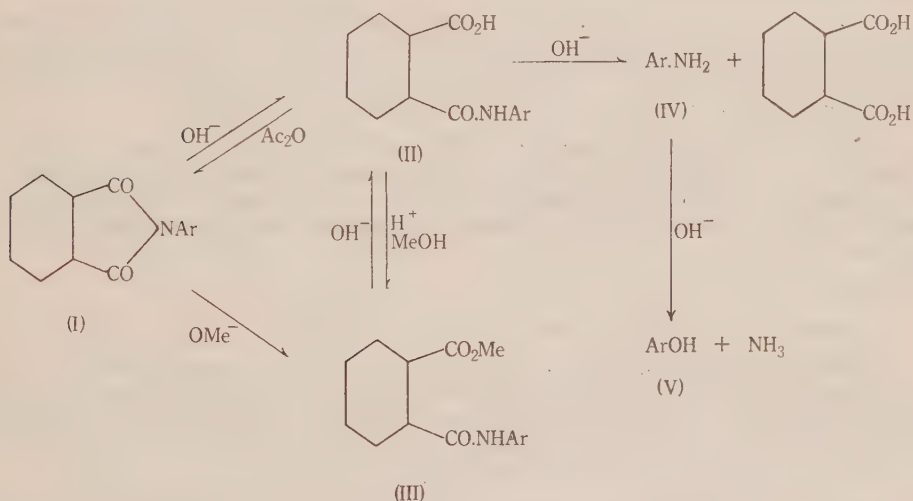
‡ Department of Chemistry, University of Western Australia, Nedlands, W.A.

or by refluxing the more reactive (Beckwith, Miller, and Leahy 1952) 1-fluoro-2,4-dinitrobenzene with sodium phthalimide in acetone for 30 minutes. The identity of the aryl phthalimide was confirmed by analysis.

The powerful $-T$ effect of the dinitrophenyl group facilitates the attack of nucleophilic reagents on the carbonyl groups of the aryl phthalimide. However, the ease of attack on the first and second CO groups is sufficiently different for intermediate compounds such as 2,4-dinitrophenylphthalamic acid and its methyl ester to be isolated by reaction with OH^- in H_2O and OMe^- in MeOH respectively. Further hydrolysis gives 2,4-dinitroaniline and phthalic acid as well as some 2,4-dinitrophenol by further reaction of the amine.

The reaction at the phthalamic acid stage may be reversed by heating with acetic anhydride, the aryl phthalimide being reformed. Also by standard procedures the phthalamic acid and its methyl ester are readily interconverted.

The reactions may be summarized as follows ($\text{Ar}=2,4$ -dinitrophenyl):



Reactions in which intermediates were not isolated are not separately shown, for example, $\text{I} \rightarrow \text{IV}$.

The obviously low reactivity of the phthalimide anion seems to be due both to a geometrical factor, shown readily by models, and the fact that the unshared N -electrons are conjugated with two CO groups. It is relevant that Brady, Quick, and Welling (1925) showed that for electrophilic aromatic substitution the phthalimide group is less o - p directing than the Me group.

Experimental

All melting points are corrected. Microanalyses by Dr. Weiler and Dr. Strauss, Oxford.

(i) *N*-(2,4-Dinitrophenyl)phthalimide (*I*).—(a) 3.4 g. of 1 fluoro-2,4-dinitrobenzene and 4 g. of potassium phthalimide were refluxed in 40 ml. dry acetone for 30 minutes, during which a deep red solution formed. Excess dilute HCl was then added. The yellow crystalline precipitate

recrystallized from acetone-water as very pale yellow leaflets, m.p. 196 °C. It recrystallizes from benzene and toluene as prisms. Yield 5 g. (87%).

Found: C, 53.9; H, 2.2; N, 13.4%.

Calculated for $C_{14}H_7O_6N_3$: C, 53.7; H, 2.2; N, 13.4%.

After completion of this work it was discovered that *N*-(2,4-dinitrophenyl)phthalimide had been reported in a note by Hodgson and Dodgson (1948), who prepared it by interacting 2,4-dinitroaniline and phthalic anhydride at high temperature. These authors do not quote a yield and give the m.p. as 192 °C. and N=13.6%.

(b) 8.4 g. of recrystallized 1-chloro-2,4-dinitrobenzene was intimately mixed with 6.0 g. of potassium phthalimide and heated in an oil-bath at 200–210 °C. for 2 hours. The resultant yellow melt was cooled and broken up under 30 ml. of hot water. The powdery solid was washed with small portions of hot water and methanol. Yield 9.2 g. (91%), m.p. 195 °C. One recrystallization from acetone-water brought the m.p. to 196 °C. (mixed m.p. with product from (a)=196 °C.).

The reaction with equivalent quantities of K_2CO_3 and phthalimide also gave the required product but in smaller yield.

(c) 300 mg. of II was mixed with 3 ml. Ac_2O and refluxed for 10 minutes. The excess anhydride was decomposed by shaking with water. After further washing 280 mg. of product (a theoretical yield), m.p. 195 °C., was obtained. On recrystallization from benzene-ethanol it formed leaflets, m.p. 196 °C. (mixed m.p. with product from (a)=196 °C.).

(ii) *N*-(2,4-Dinitrophenyl)phthalamic Acid (II).—(a) 3.6 g. of I was dissolved in 40 ml. acetone and 50 ml. 0.28N NaOH (aq.) added. After 3 minutes at room temperature the mixture was acidified with dilute HCl and diluted to 250 ml. The product, m.p. 157 °C., was practically pure. Yield 3.7 g. (or theoretical). It recrystallized from benzene as very pale yellow prisms, m.p. 158.5 °C.

Found: C, 50.6; H, 2.6; N, 12.4%.

Calculated for $C_{14}H_9O_7N_3$: C, 50.8; H, 2.7; N, 12.7%.

(b) 600 mg. of III was dissolved in 10 ml. acetone and 10 ml. 0.28N NaOH (aq.) added. A red solution formed which after standing for 2 minutes was acidified with dilute HCl. The colour disappeared and the product on pouring into water gave crystals which were filtered and washed. After recrystallization from benzene, 500 mg. (89% yield) of acid, m.p. 158.5 °C., was obtained (mixed m.p. with product from (a)=158.5 °C.).

(iii) *Methyl N*-(2,4-Dinitrophenyl)phthalamate (III).—(a) 2 g. of I was dissolved in a mixture of 20 ml. acetone, and 20 ml. methanol and 4 ml. 2.8N sodium methoxide (in methanol) added dropwise. The solution became red at once and after 2 minutes was neutralized with dilute HCl, and then diluted with 40 ml. water. The ester formed a copious white precipitate, m.p. 162 °C. Yield 2.1 g. (95%).

Found: C, 52.4; H, 3.3; N, 11.9%.

Calculated for $C_{15}H_{11}O_7N_3$: C, 52.2; H, 3.2; N, 12.2%.

(b) 400 mg. of II was mixed with 8 ml. of dry methanol and 0.4 ml. of conc. H_2SO_4 and refluxed for 1 hour. On cooling the yellow solution a dense crystalline precipitate formed. After recrystallization from methanol it gave 220 mg. (53%) of ester, m.p. 162 °C. (mixed m.p. with product from (a) 162 °C.).

(iv) *Hydrolyses to 2,4-Dinitroaniline (IV)*.—(a) 0.2 g. of I was refluxed with 4 ml. of 2% NaOH for 5 minutes by which time all the solid had dissolved. On cooling a yellow solid was precipitated. This was collected and recrystallized from aqueous ethanol, m.p. 181 °C. This was shown to be 2,4-dinitroaniline by mixed m.p. with an authentic sample. The filtrate on acidification gave a small amount of II, m.p. 158 °C.

(b) 120 mg. of II was refluxed with 4 ml. 2% NaOH. A red solution formed and after 2 minutes a yellow precipitate. The smell of ammonia was apparent after 5 minutes and heating was stopped. The mixture was cooled and the yellow precipitate identified as 2,4-dinitroaniline. Yield 50 mg. (75%).

(c) A similar procedure with 200 mg. of III gave 85 mg. of amine (90%) and some phthalamic acid.

The authors wish to acknowledge financial assistance from the Research Grant to Australian Universities.

References

- BECKWITH, A. L., MILLER, J., (and in part) LEAHY, G. D. (1952).—*J. Chem. Soc.* **1952**: 3552.
BRADY, O. L., QUICK, W. G. E., and WELLING, W. F. (1925).—*J. Chem. Soc.* **127**: 2265.
HODGSON, H. H., and DODGSON, D. P. (1948).—*J. Chem. Soc.* **1948**: 1995.

CORRIGENDA

VOLUME 5, NUMBER 2

Page 357, equation (22): For $\frac{N'+1}{m+\frac{1}{2}}$ read $\frac{m+\frac{1}{2}}{N'+1}$.

Page 361, line 9 from bottom: For 0.1 g. of each solvent read 0.1 g. of each solute.

Page 363, Table .1, (n_1+n_2) values of maleic acid: For 7+15 read 7+8.

INDEX

	PAGE		PAGE
Abbey, R. L.—		Amines, <i>N</i> -Heterocyclic, Di-	
The Velocity of Sound in		sulphonyl Derivatives of ..	368
Gases at Low Pressures	223	Amines, The Reaction of	
Absorption and Reflection of		<i>N</i> -Heterocyclic, with Sul-	
Microwave Radiation by a		phonyl Halides	374
Mercury-Vapour Discharge	592	Amino Acids and Proteins, The	
Absorption, Dielectric, and D.C.		Reaction of, with Diazonium	
Conductivity in <i>n</i> -Primary		Compounds	736
Alcohols	671	Amino Acids, Preparation of	
Acylamino Acids, Anhydride		Some 2-Thiohydantoins from	711
Intermediates in the Forma-		Amperometric and Starch In-	
tion of 1 - Acyl - 2 -		dicator Methods for the	
thiohydantoins from ..	728	Titration of Iodine and	
Acylamino Acids, 2 - Thio-		Thiosulphate, An Investi-	
hydantoins from	711	gation of Errors in the ..	541
Adams, C. S., and Mellor, D. P.—		Anet, F. A. L., Gilham, P. T.,	
The Crystal Structure of		Gow, P., Hughes, G. K.,	
Potassium Hexachloro-		and Ritchie, E.—	
ruthenate (IV)	577	The Chemical Constituents of	
Adsorption at Liquid Surfaces,		Australian <i>Flindersia</i>	
The Kinetics of	683	Species. III. The Alkaloids	
Adsorption on Solid Surfaces,		of <i>Flindersia collina</i> Bail... ..	412
The Hysteresis of	288	Angyal, S. J.—	
Admittance, Microwave, of a		Sulphonamides. IV. The	
Mercury-Vapour Discharge	607	Reaction of <i>N</i> -Heterocyclic	
Air-Water Interface, Rigidity of		Amines with Sulphonyl	
Adsorbed Films at an ..	189	Halides	374
Alcohols, Long-Chain Aliphatic,		Angyal, S. J., Morris, W. O.,	
Temperature Dependence of		and Warburton, W. K.—	
the Dielectric Properties of..	661	Sulphonamides. III. Disul-	
Alcohols, <i>n</i> -Primary, Dielectric		phonyl Derivatives of some	
Absorption and D.C. Con-		<i>N</i> -Heterocyclic Amines ..	368
ductivity in	671	Anisotropy of Electrical Re-	
Aldehydes, Lignin	346	sistivity of Deformed Cubic	
Alexander, A. E.— <i>See</i> Cumper,		Metals and Alloys ..	128
C. W. N.	146, 153, 189	Anisotropy of Electrical Re-	
Alkaloids, Coloured, The		sistivity of Some Cubic	
Structures of	406	Metals and Alloys	119
Alkaloids of the Australian		Aromadendrin, Kaempferol, and	
Rutaceae	387, 401,	Ellagic Acid	379
	406, 563, 768	Ashelford, N., and Mellor, D. P.—	
Alloys and Cubic Metals, De-		Attempts to Effect Exchange	
formed, Electrical Resistivity		and Displacement Re-	
of	128	actions with Cobalt (II)	
Alloys and Cubic Metals,		Mesoporphyrin	784
Electrical Resistivity of ..	119	Aspartic Acid, Thiohydantoins	
		Derived from	721

	PAGE		PAGE
Atebrin, The Production of Intermediates in the Manufacture of	198	Brown, R. D.— Conjugation Energy	339
Australian Plants, Colouring Matters of	760	Burbury, D. W. P., and Fenton, K. B.— The High Latitude East-West Asymmetry of Cosmic Rays	47
Australian Rutaceae, Alkaloids of	387, 401, 406, 563, 768	Burbury, D. W. P., and McLaren, A. C.— Cosmic Ray Asymmetry Studies at Hobart, Tasmania	782
Azo-Derivatives of Histidine and Tyrosine, A Spectrophotometric Study of	736	Cannon, J. R., Hughes, G. K., Neill, K. G., and Ritchie, E.— Alkaloids of the Australian Rutaceae: <i>Erodia xanthoxyloides</i> F. Muell. III. The Structures of the Coloured Alkaloids Evonxanthidine, Xanthevodine, and Xanthoxoline	406
Beckwith, A. L., and Miller, J.— On <i>N</i> - (2,4 - Dinitrophenyl)-phthalimide	786	Cannon, J. R., Hughes, G. K., Price, J. R., and Ritchie, E.— The Chemical Constituents of Australian <i>Flindersia</i> Species. IV. The Constituents of <i>Flindersia bourjotiana</i> F. Muell.	420
Bending of a Semi-infinite Strip	227	Canthin-6-one, Identification of	387
Benzene- <i>n</i> -Heptane System	530	¹⁴ C-Formaldehyde with Dime-done, Isotope Effect in the Reaction of	521
Bland, D. E., Hillis, W. E., and Williams, E. J.— Prediction Curves for Counter-Current Separations: Application to Lignin Aldehydes	346	Chapman, Sydney— The Calculation of the Probable Error of Determinations of the Lunar Daily Harmonic Component Variations in Geophysical Data: A Correction	218
Bradbury, J. H., and Hambly, A. N.— An Investigation of Errors in the Amperometric and Starch Indicator Methods for the Titration of Millinormal Solutions of Iodine and Thiosulphate	541	Christiansen, W. N., and Hindman, J. V.— A Preliminary Survey of 1420 Mc s. Line Emission from Galactic Hydrogen	437
Brass, ($\alpha + \beta$), Deformation of	114	Clarebrough, L. M., and Perger, G. R.— Influence of the Volume Fractions of the Phases on the Deformation of ($\alpha + \beta$) Brass	114
Breyer, B., and Hacobian, S.— Tensammetry: A Method of Investigating Surface Phenomena by A.C. Current Measurements	500	Clothier, W. K.—See Broom, T.	119
Broom, T.— On the Anisotropy of Electrical Resistivity of Deformed Cubic Metals and Alloys	128		
Broom, T., and Clothier, W. K.— The Anisotropy of Electrical Resistivity of Cold Drawn Wires of Some Cubic Metals and Alloys	119		
Brown, I.— Liquid - Vapour Equilibria. III. The Systems Benzene- <i>n</i> - Heptane, <i>n</i> - Hexane - Chlorobenzene, and <i>cyclo</i> -Hexane-Nitrobenzene	530		

	PAGE		PAGE
Cloud Drops, The Growth of	59, 473	Davies, W., Olver, N. H., and Wilson, B. W.—	
Cobalt(II) Mesoporphyrin, Re- actions with	784	The Production of Inter- mediates in the Manu- facture of Atebrin.. ..	198
Colouring Matters from Aust- ralian Plants	760	Demethylation of Methoxy- acridones	206
Condensation of a Vapour at a Crystalline Surface	628	Diazonium Compounds, The Reaction of Amino Acids and Proteins with	736
Condensation, The Growth of Cloud Drops by	59, 473	Dielectric Absorption in <i>n</i> - Primary Alcohols	671
Conductivity, D.C., and Di- electric Absorption in <i>n</i> - Primary Alcohols	671	Dielectric Properties of Dipolar Substances, The Frequency Dependence of the	647
Conjugation Energy	339	Dielectric Properties of Long- Chain Aliphatic Alcohols in the Solid State	661
Cooke, R. G., and Dowd, H.—		Dielectric Properties of Straight Polar Chains	135
Colouring Matters of Aust- ralian Plants. II. Naphtho- quinones from <i>Diospyros</i> <i>hebecarpa</i> A. Cunn.	760	Dimedone, Isotope Effect in the Reaction of ¹⁴ C-Formal- dehyde with	521
Corrigenda	789	<i>N</i> - (2,4 - Dinitrophenyl)phth- alimide	786
Cosmic Radio Radiation, The Distribution of the Discrete Sources of	266	<i>Diospyros hebecarpa</i> , Naphtho- quinones from	760
Cosmic Radio Radiation, The Positions of Six Discrete Sources of	456	Disulphonyl Derivatives of Some <i>N</i> -Heterocyclic Amines	368
Cosmic Ray Asymmetry Studies at Hobart, Tasmania	782	Dowd, H.— <i>See</i> Cooke, R. G...	760
Cosmic Rays, The High- Latitude East - West Asymmetry of	47	Downes, A. M.—	
Counter-Current Separation, Prediction in	346	Isotope Effect in the Reaction of ¹⁴ C-Formaldehyde with Dimedone	521
Crystal Structure of Potassium Hexachlororuthenate(IV) ..	577	Dryden, J. S.—	
Cumper, C. W. N., and Alexander, A. E.—		Temperature Dependence of the Dielectric Properties of Long-Chain Aliphatic Al- cohols in the Solid State..	661
The Viscosity and Rigidity of Gelatin in Concentrated Aqueous Systems.		Dyer, A. J.—	
I. Viscosity	146	The Stopping Power of Ilford C ₂ Emulsion	104
II. Rigidity	153		
The Surface Chemistry of Proteins. III. The Rigidity of Adsorbed Films at an Air-Water Interface	189	Electromagnetic Equations, A Simplified Form of the Relativistic	423
Curve-Fitting by the Method of Grouping	238	Electron Beam, The Optimum Space - Charge - Controlled Focus of an	430
Cygnus, Constellation of, Radio-Frequency Radiation from the	17	Electrons, Fast, Scattering of..	258
		Ellagic Acid	379

	PAGE		PAGE
Emulsion, Ilford C ₂ , The Stopping Power of ..	104	Gow, P.—See Anet, F. A. L., et al. ..	412
Evaporation from Shallow Ponds, The Rate of Natural ..	315	Guest, P. G.— Curve-Fitting by the Method of Grouping ..	238
<i>Evodia xanthoxyloides</i> , Alkaloids of ..	401, 406	Gunnarsen, E. M.— The Scattering of Fast Electrons and Positrons by Heavy Elements ..	258
Evoxanthidine ..	406		
Ewers, W. E., and Sutherland, K. L.— The Rôle of Surface Transport in the Stability and Breakdown of Foams ..	697	Hacobian, S.—See Breyer, B. . .	500
Fenton, K. B.—See Burbury, D. W. P. ..	47	Haemins, Solid, The Magnetic Properties and Chemical Structures of ..	173
Ferguson, J.— The Rate of Natural Evaporation from Shallow Ponds ..	315	Halides, Sulphonyl, The Reaction of <i>N</i> -Heterocyclic Amines with ..	374
Films, Adsorbed, Rigidity at an Air-Water Interface ..	189	Hambly, A. N.—See Bradbury, J. H. ..	541
Flavones, New, from <i>Pongamia pinnata</i> ..	754	Hamon, B. V., and Meakins, R. J.— Dielectric Absorption and D.C. Conductivity in <i>n</i> -Primary Alcohols ..	671
Foams, The Rôle of Surface Transport in the Stability and Breakdown of ..	697	Haynes, H. F., Nelson, Eva R., and Price, J. R.— Alkaloids of the Australian Rutaceae: <i>Pentaceras australis</i> Hook. F. I. Isolation of the Alkaloids and Identification of Canthin-6-one ..	387
<i>Flindersia bourjotiana</i> ..	420	Heating, Surface, Approximations in Transient ..	1
<i>Flindersia collina</i> ..	412	Hexachlororuthenate (IV), Potassium. The Crystal Structure of ..	577
Fraser, Dallas.—See Higgins, H. G. ..	736	<i>cyclo</i> Hexane — Nitrobenzene System ..	530
Galaetic Hydrogen, A Preliminary Survey of 1420 Mc/s. Line Emission from ..	437	<i>n</i> -Hexane — Chlorobenzene System ..	530
Gelatin, Viscosity and Rigidity of ..	146, 153	Heymann, E., the Late.—See Gilby, A. R. ..	160
Geophysical Data, Lunar Daily Harmonic Component Variations in ..	218	Higgins, H. G., and Fraser, Dallas— The Reaction of Amino Acids and Proteins with Diazonium Compounds. I. A Spectrophotometric Study of Azo-Derivatives of Histidine and Tyrosine ..	736
Gilby, A. R., and Heymann, E., the Late— Oleic Acid Monolayers on Concentrated Salt Solutions ..	160		
Gilham, P. T. — See Anet, F. A. L., et al. ..	412		
Gipps, G. de V.—See Strohhfeldt, M., and McNicol, R. W. E. . .	464		
Glutamic Acid, Thiohydantoin Derived from ..	721		
<i>Glycosmis pentaphylla</i> , Alkaloids of ..	579		

	PAGE		PAGE
Higgins, H. G., and Williams, E. J.—		Iodine and Thiosulphate, Millinormal Solutions of, An Investigation of Errors in ..	541
Kinetic Treatment Appropriate to Rapid Reactions involving Two Consecutive Second Order Steps	572	Ionospheric Measurements at Oblique Incidence over Eastern Australia	464
Hillis, W. E.—		Isotope Effect in the Reaction of ^{14}C -Formaldehyde with Dimedone	521
The Chemistry of Eucalypt Kinols. II. Aromadendrin, Kaempferol, and Ellagic Acid	379	Jaeger, J. C.—	
Hillis, W. E.—See Bland, D. E., and Williams, E. J. ..	346	Approximations in Transient Surface Heating	1
Hindman, J. V. — See Christiansen, W. N. ..	437	Kaempferol	379
Hirst, F.—		Keratin, Urea Solutions of, The Viscosity of	555
Angular Distribution of α -Particles from the $\text{Li}^7(p,\alpha)\text{He}^4$ Reaction at Low Proton Bombarding Energies	570	Kinetic Treatment involving Two Consecutive Steps ..	572
Histidine and Tyrosine, A Spectrophotometric Study of Azo-Derivatives of	736	Kinetics of Adsorption at Liquid Surfaces	683
Hollway, D. L.—		Kinos, Eucalypt, The Chemistry of	379
The Optimum Space-Charge-Controlled Focus of an Electron Beam	430	Lignin Aldehydes	346
Hughes, G. K., Matheson, N. K., Norman, A. T., and Ritchie, E.—		Liquid Phase Embryos at the Intersections of Plane Solid Surfaces	618
The Demethylation of Methoxyacridones	206	Liquid-Vapour Equilibria ..	530
Hughes, G. K., Neill, K. G., and Ritchie, E.—		$\text{Li}^7(p,\alpha)\text{He}^4$ Reaction at Low Proton Bombarding Energies	570
Alkaloids of the Australian Rutaceae: <i>Evodia xanthoxyloides</i> F. Muell. II. Isolation of the Alkaloids from the Leaves	401	Little, A. G.—See Payne-Scott, Ruby	32
Hughes, G. K.—		Lunar Variations in Geophysical Data	218
See Anet, F. A. L., <i>et al.</i> ..	412	McKenzie, A. W., and Price, J. R.—	
See Cannon, J. R., <i>et al.</i> ..	406, 420	Alkaloids of the Australian Rutaceae: <i>Glycosmis pentaphylla</i> (Retz.) Correa ..	579
Huxley, L. G. H.—		McLaren, A. C.—See Burbury, D. W. P.	782
The Persistence of Meteor Trails	10	McNicol, R. W. E. — See Strohfeldt, M., and Gipps, G. de V.	464
Hysteresis of Adsorption on Solid Surfaces	288	Mansfield, W. W.—	
		The Spontaneous Emulsification of Mixtures of Oleic Acid and Paraffin Oil in Alkaline Solutions ..	331

	PAGE		PAGE
Matheson, N. K.— <i>See</i> Hughes, G. K., <i>et al.</i>	206	Nelson, Eva R., and Price, J. R.—	
Meakins, R. J.— <i>See</i> Hamon, B. V.	671	Alkaloids of the Australian Rutaceae: <i>Pentaceras australis</i> Hook. F.	
Mellor, D. P.—		II. Identification of 5-Methoxycanthinone ..	563
<i>See</i> Adams, C. S.	577	III. Identification of 4-Methylthiocanthinone	768
<i>See</i> Ashelford, N.	784	Norman, A. T.— <i>See</i> Hughes, G. K., <i>et al.</i>	206
Mercury-Vapour Discharge, The Absorption and Reflection of Microwave Radiation by a ..	592	O'Dwyer, J. J., and Sack, R. A.—	
Mercury-Vapour Discharge, The Microwave Admittance of a ..	607	The Frequency Dependence of the Dielectric Properties of Dipolar Substances ..	647
Mesoporphyrin, Cobalt(II), Reactions with	784	Oleic Acid in Alkaline Solutions	331
Metals, Cubic, Electrical Resistivity of	119	Oleic Acid Monolayers on Salt Solutions	160
Metals, Cubic, Electrical Resistivity of Deformed ..	128	Olver, N. H.— <i>See</i> Davies, W., and Wilson, B. W. ..	198
Meteor Trails, The Persistence of	10	Paraffin Oil in Alkaline Solutions	331
Methoxyacridones. Demethylation of	206	α -Particles from the $\text{Li}^7(p,\alpha)\text{He}^4$ Reaction at Low Proton Bombarding Energies ..	570
5-Methoxycanthinone, Identification of	563	Payne-Scott, Ruby, and Little, A. G.—	
4-Methylthiocanthin - 6 - one, Identification of	768	The Position and Movement on the Solar Disk of Sources of Radiation at a Frequency of 97 Mc/s. III. Outbursts	32
Miller, J.— <i>See</i> Beckwith, A. L.	786	<i>Pentaceras australis</i> , Alkaloids of	387, 563, 768
Mills, B. Y.—		Perger, G. R.— <i>See</i> Clarebrough, L. M.	114
The Distribution of the Discrete Sources of Cosmic Radio Radiation ..	266	Piddington, J. H., and Minnett, H. C.—	
The Positions of Six Discrete Sources of Cosmic Radio Radiation	456	Radio-Frequency Radiation from the Constellation of Cygnus	17
Minnett, H. C.— <i>See</i> Piddington, J. H.	17	<i>Pongamia pinnata</i> , New Flavones from	754
Monolayers, Oleic Acid, on Salt Solutions	160	Positrons, Scattering of, by Heavy Elements	258
Morris, W. O.— <i>See</i> Angyal, S. J., and Warburton, W. K.	368	Price, J. R.—	
Naphthoquinones from <i>Diospyros hebecarpa</i> ..	760	<i>See</i> Cannon, J. R., <i>et al.</i> ..	420
Neill, K. G.—		<i>See</i> Haynes, H. F., and Nelson, Eva R.	387
<i>See</i> Cannon, J. R., <i>et al.</i> ..	406	<i>See</i> McKenzie, A. W.	579
<i>See</i> Hughes, G. K., and Ritchie, E.	401	<i>See</i> Nelson, Eva R. ..	563, 768
Nelson, Eva R.— <i>See</i> Haynes, H. F., and Price, J. R. ..	387		

	PAGE		PAGE
Prime, H. A.—		Row, L. Ramachandra—	
The Absorption and Reflec-		New Flavones from <i>Pongamia</i>	
tion of Microwave Radia-		<i>pinnata</i> (L.) Merr. ..	754
tion by a Mercury-Vapour		Sack, R. A.—	
Discharge	592	The Dielectric Properties of	
The Microwave Admittance of		Systems Containing Straight	
a Mercury-Vapour Dis-		Polar Chains	135
charge	607	<i>See</i> O'Dwyer, J. J. ..	647
Proteins, The Reaction of, with		Scutt, P. B.— <i>See</i> Rawlinson,	
Diazonium Compounds ..	736	W. A.	173
Proteins, The Surface Chemistry		Smith, E. J.—	
of	189	The Influence of Meteoro-	
Proton Bombarding Energies,		logical Conditions on Rain-	
Angular Distribution of		making in the Sydney Area	87
α -Particles at Low	572	Smith, R. C. T.—	
Radiation at 97 Mc/s. ..	32	The Bending of a Semi-	
Radiation, Cosmic Radio, The		infinite Strip	227
Discrete Source of	266	Soil Temperature Wave, Annual	303
Radiation, Cosmic Radio, The		Solar Disk, The Position of	
Positions of Six Discrete		Sources of Radiation at	
Sources of	456	97 Mc/s. on the	32
Radiation, High-Energy, The		Sound, The Velocity of, in	
Absolute Measurement of ..	581	Gases at Low Pressures ..	223
Radiation, Microwave.. ..	592	Spicer, B. M.—	
Radiation, Movements of		The Absolute Measurement	
Sources of, on the Solar Disk	32	of High-Energy Radiation	581
Radiation, Radio-Frequency,		Squires, P.—	
from the Constellation of		The Growth of Cloud Drops	
Cygnus	17	by Condensation.	
Rainmaking, The Influence of		I. General Characteristics	59
Meteorological Conditions on	87	II. The Formation of	
Rawlinson, W. A., and Scutt,		Large Cloud Drops ..	473
P. B.—		Strohfeldt, M., McNicol,	
The Magnetic Properties and		R. W. E., and Gipps,	
Chemical Structures of		G. de V.—	
Solid Haemins	173	Ionospheric Measurements at	
Resistivity, Electrical, of De-		Oblique Incidence over	
formed Cubic Metals and		Eastern Australia.. ..	464
Alloys	128	Sulphonamides	368, 374
Resistivity, Electrical, of Some		Sulphonyl Halides	374
Cubic Metals and Alloys ..	119	Sutherland, K. L.—	
Rigidity of Adsorbed Films at		The Kinetics of Adsorption	
an Air-Water Interface ..	189	in Liquid Surfaces.. ..	683
Rigidity of Gelatin	153	<i>See</i> Ewers, W. E.	697
Ritchie, E.—		Swan, J. M.—	
<i>See</i> Anet, F. A. L., <i>et al.</i> ..	412	Thiohydantoins.	
<i>See</i> Cannon, J. R., <i>et al.</i> 406,	420	I. The Preparation of some	
<i>See</i> Hughes, G. K., <i>et al.</i> ..	206	2-Thiohydantoins from	
<i>See</i> Hughes, G. K., and		Amino Acids and Acyl-	
Neill, K. G.	401	amino Acids	711
		II. Thiohydantoins De-	
		rived from Aspartic and	
		Glutamic Acids.. ..	721

	PAGE		PAGE
III. Anhydride Inter-		Viscosity of Urea Solutions of	
mediates in the Forma-		Keratin	555
tion of 1-Acyl-2-thio-		Warburton, W. K.— <i>See</i> Angyal,	
hydantoins from Acyl-		S. J., and Morris, W. O. ..	368
amino Acids	728	West, E. S.—	
Taylor, N. W.—		A Study of the Annual Soil	
A Simplified Form of the		Temperature Wave ..	303
Relativistic Electromag-		Williams, E. J.—	
netic Equations	423	<i>See</i> Bland, D. E., and Hillis,	
Tensammetry: A Method of		W. E.	346
Investigating Surface Phen-		<i>See</i> Higgins, H. G.	572
omena by A.C. Measurements	500	Wilson, B. W.— <i>See</i> Davies, W.,	
Thiohydantoins	711, 721, 728	and Olver, N. H.	198
Thiosulphate and Iodine, Milli-		Woods, E. F.—	
normal Solutions of, An		The Viscosity of Urea Solu-	
Investigation of Errors in		tions of Keratin	555
Titration of	541	Wylie, R. G.—	
Tyrosine and Histidine, A		On the Hysteresis of Adsorp-	
Spectrophotometric Study of		tion on Solid Surfaces ..	288
Azo-Derivatives of	736	The Condensation of a Vapour	
Urea Solutions of Keratin, The		at a Crystalline Surface ..	628
Viscosity of	555	The Properties of Liquid	
Vapour, Condensation of a, at		Phase Embryos at the	
a Crystalline Surface ..	628	Intersections of Plane Solid	
Viscosity of Gelatin	146	Surfaces	618
		Xanthevodine and Xanthoxoline	406

DATE DUE

REC'D PER

AUG 20 1983

RETD AUG 19 1983

DEC 17 '84 PER

RETD DEC 17 1984

UIC Rec'd OCT 25 2012

PERIODICALS MUST BE RETURNED
TO PERIODICALS DESK ONLY



3 8198 304 260 639

UNIVERSITY OF ILLINOIS AT CHICAGO

